Characterising new roles for APOBEC4 and ADAR Deaminases

Marion Hogg

Thesis presented for the degree of Doctor of Philosophy,

The University of Edinburgh 2010.

Abstract

Deamination or the hydrolytic removal of one hydroxyl group from a base in DNA or RNA can lead to changes in the transcript and protein produced. Examples of this are the deamination of cytosine residues in DNA by activation induced deaminase (AID) during antibody diversification, or deamination of adenosine at the Q/R site in the GluR-B transcript by adenosine deaminase acting on RNA 2 (ADAR2), which regulates calcium permeability in neurons. The initial focus of my thesis was to characterise a putative novel deaminase APOBEC4. APOBEC4 was identified in a bioinformatic search for proteins containing the core catalytic residues common to the whole family of Cytidine Deaminase enzymes. The aim of the project was to express and purify recombinant APOBEC4 for *in vitro* characterization, however despite using different expression systems and purification conditions the majority of the recombinant protein was inherently insoluble and I could not isolate sufficient amounts of protein for further studies. Recombinant protein with a GST-tag was used to generate polyclonal antibodies which recognised recombinant protein but were unable to detect endogenous APOBEC4.

The focus of my thesis then changed to the process of adenosine to inosine editing in RNA, which is a post-transcriptional mechanism for generating protein diversity. The enzyme family responsible for catalysing this reaction is known as ADAR, and *Drosophila melanogaster* has only one *Adar* gene. Flies lacking the *Adar* gene show locomotion defects and age-dependent neurodegeneration, however little is known about the molecular mechanism underlying these defects. To investigate this phenotype I performed microarray analysis on RNA isolated from heads of 5 day old flies lacking the *Adar* gene to characterize gene expression changes in the fly heads before neurodegeneration caused secondary effects. Analysis was also performed on *Adar*-null flies expressing either an active *Adar* gene or a catalytically-inactive *Adar* gene in cholinergic neurons to determine which transcripts could be directly regulated by *Adar*. I confirmed the microarray results by real-time PCR, and demonstrated that the changes in transcript level could be reversed by expression of either active or catalytically-inactive *Adar*. Expression of edited transcripts did not

change dramatically. Filter-binding analysis and electrophoretic mobility shift assay revealed that recombinant ADAR could bind to all RNA transcripts analysed with similar affinity; both known substrates and potential new substrates for *Adar*, as well as transcripts that were chosen as negative controls due to their expression not altering in the expression microarray. Recombinant ADAR bound to dsRNA with a very high affinity; other transcripts investigated bound with considerably lower affinity, yet all transcripts investigated were bound by ADAR.

Further analysis of transcript changes in *Adar*-null flies was investigated by performing microarray analysis with a custom-made splicing-sensitive microarray. Analysis revealed that a subset of transcripts were differentially spliced in *Adar*-null flies; however this group of transcripts was distinct from the group identified as being altered on the expression microarray, indicating that the splicing changes are independent of changes in expression. Analysis of exon-specific probes on the splicing array confirmed the transcript changes identified in the expression array. Real-time PCR confirmed the changes in splicing, and these transcripts were further examined by sequence analysis. This revealed several transcripts identified as altered by the AS array showed use of alternative polyadenylation sites indicating ADAR may have a role in detemining polyadenylation site selection.

Declaration

I declare that the work presented in this PhD is my own and has not been submitted for any other degree, unless otherwise stated.

Marion Hogg
June 2010

Acknowledgements

I would like to thank Mary for the opportunity to work in her lab and her support over the years.

Thanks to Leeanne for inspiring me to see this PhD through to the end…
And of course a huge amount of thanks to Jim for helping me with site-directed mutagenesis and the French press and everything else that he has helped with over the years. I would also like to thank Bret for his unwavering optimism and his support over the years. Cheers to the Italians for introducing the lab gnocchi!

A big thanks to Ian Adams for all his help and guidance with the APOBEC4 project, one day maybe someone will figure it out!

Thanks to Marco Blanchette and Graeme Grimes for helping me with microarray analysis and explaining the language of R. Thanks to Paul Perry for help with imaging and Craig Nicol for helping with posters, Adobe Illustrator, and printing this huge thesis.

Thanks to all the technical services staff for their help over the years but in particular Stewart McKay, Agnes Gallagher and Stephen Brown for helping me sequence thousands of clones in the hope of finding some editing!

Thanks to the Caceres lab post-docs for imparting some of their extensive knowledge of RNA! And last but not least thanks to the other HGU students for the pantomimes, pirate parties and of course the legendary peebles retreat!

A special mention has to go to my parents for their continued support throughout the PhD, thanks for everything!

Abbreviations

aa	amino acids
ACF	APOBEC Complementation Factor
Adar	Adenosine deaminase that acts on RNA
AID	Activation induced deaminase
ALS	Amyotrophic Lateral Sclerosis
Apo	Apolipoprotein
APOBEC	Apolipoprotein B mRNA-editing enzyme catalytic polypeptide
A1/2/3/4	APOBEC1/2/3/4
AS	Alternative splicing
Bp	base pairs
BP	Branch-Point
CSR	Class switch recombination
CNS	Central Nervous System
^{o}C	degrees Celsius
dADAR	*Drosophila melanogaster* ADAR
DAPI	4', 6-diamidino-2-phenylindole
dH_2O	distilled water
DNA	Deoxyribonucleic Acid
dpp	days post partum
dsDNA	double-strand DNA
dsRBD	double-stranded RNA binding domain
dsRNA	double-strand RNA
EDTA	Ethylene diamine tetra-acetic acid
EMSA	Electrophoretic Mobility Shift Assay
ESE	Exonic Splicing Enhancer

ESS	Exonic Splicing Silencer
EST	Expressed Sequence Tag
FCS	Foetal Calf Serum
FFV	Feline foamy virus
FITC	Fluorescein isothiocyanate
FLIS6	Amino-terminal FLAG and carboxy-terminal 6x histidine epitope tags
G	G centrifugal force
g	gram
GEO	Gene Expression Omnibus
GST	Glutathione S-transferase
H_2O	water
HDL	High Density Lipoprotein
HERV	Human Endogenous Retrovirus
IP	Immunoprecipitation
IPTG	
IRES	Internal Ribosome Entry Site
ISE	Intronic Splicing Enhancer
ISS	Intronic Splicing Silencer
kb	kilo base
kDa	kiloDalton
L	Litre
LB	Luria Broth
LDL	Low Density lipoprotein
LINE/L1	Long interspersed repeat element/1
LTR	Long terminal repeat

M	Molar
MEFs	Mouse Embryonic Fibroblasts
ml	millilitre
mM	millimolar
µl	microlitre
mRNA	messenger RNA
ncRNA	non-coding RNA
NES	Nuclear Export Signal
NLS	Nuclear Localisation Signal
nm	nanometre
nt	nucleotide
NTA	Nitrilotriacetic Acid
NTD	Amino-terminal Domain
ORF	Open Reading Frame
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
RIN	RNA Integrity Number
RIP	Ribonucleoprotein Immuno-precipitation
RISC	RNA-induced silencing complex
RNA	Ribonucleic Acid
RNAi	RNA interference
RNAP II	RNA polymerase II
RNase	Ribonuclease
rpm	revolutions per minute
RRM	RNA Recognition Motif
rRNA	ribosomal RNA
RT	room temperature

RT-PCR	Reverse Transcriptase – PCR
S2	Schneider's S2 cell
SD	standard deviation
SDS	Sodium Dodecyl Sulphate
SDS-PAGE	SDS – PolyAcrylamide Gel Electrophoresis
Sf9	Spodoptera frugiperda 9 cells
SHM	Somatic Hypermutation
shRNAi	short hairpin RNA interference
SINES	Short interspersed repeat elements
snoRNA	small nucleolar RNA
snRNA	small nuclear RNA
snRNP	small nuclear ribonucleoprotein
ss	splice site
ssDNA	single-strand DNA
ssRNA	single-strand RNA
Sxl	Sex lethal
TDG	Thymine DNA Glycosylase
Tris	Tris(hydroxymethyl)-amino-methane
tRNA	transfer RNA
Tudor-SN	Tudor staphylococcal nuclease
UTR	Untranslated Region
UV	Ultraviolet
VLDL	Very Low Density Lipoprotein

Single Letter Amino Acid Code

A	**Alanine**	**Aliphatic**	**L**	**Leucine**	**Aliphatic**
R	**Arginine**	**Basic**	**K**	**Lysine**	**Basic**
N	**Aspargine**	**Acidic**	**M**	**Methionine**	**Sulphur containing**
D	**Aspartate**	**Acidic**	**F**	**Phenylalanine**	**Aromatic**
C	**Cysteine**	**Sulphur containing**	**P**	**Proline**	**Imino**
Q	**Glutamine**	**Acidic**	**S**	**Serine**	**Non-Aromatic**
E	**Glutamate**	**Acidic**	**T**	**Threonine**	**Non-Aromatic**
G	**Glycine**	**Aliphatic**	**W**	**Tryptophan**	**Aromatic**
H	**Histidine**	**Basic**	**Y**	**Tyrosine**	**Aromatic**
I	**Isoleucine**	**Aliphatic**	**V**	**Valine**	**Aliphatic**

Table of contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1 Editing

The term "RNA Editing" was originally coined to describe the insertion of uridine residues in trypanosome mitochondrial cytochrome oxidase subunit II (*CoxII*) mRNA (Benne *et al.*, 1986). Since then the term has been used to describe insertion or deletion of nucleotides, as well as conversion of one nucleotide to another. Uridine to cytidine (U-to-C) RNA editing, or amination, has been observed in the RNA for the Wilms' Tumour susceptibility gene (WT1) in mammals (Sharma *et al.*, 1994). However this thesis will focus on the process of base conversion editing. Hydrolytic deamination or removal of an amine group to release ammonia can occur on several bases within DNA or RNA: adenosine to inosine (A-to-I) in RNA; cytosine to uracil (C-to-U) in DNA or RNA; and 5-methylcytosine to thymine (5MeC-to-T) in DNA, see Figure 1.1. These deamination events are mediated by cytidine deaminase (CDA) enzymes which include the adenosine deaminases that act on RNA (ADAR) family, and the apolipoprotein B mRNA-editing enzyme catalytic polypeptide (APOBEC) family.

The consequences of base conversion can be dramatic. When occurring in DNA base conversion can induce chromosomal rearrangements or loss of the epigenetic 5-methylcytosine mark. In messenger RNA (mRNA) nucleotide conversion can change the coding capacity of the transcript, alter splice donor, acceptor or regulatory sequences, or create alternative start or stop codons. Hence the need for a greater understanding of the enzymes that perform these reactions and their substrates. In this thesis I attempted to characterise a novel member of the cytidine deaminase family APOBEC4 however this proved to be technically challenging. I then went on to investigate the role of *Drosophila mealogaster* ADAR as an RNA-binding protein, which is independent of its role as an adenosine deaminase.

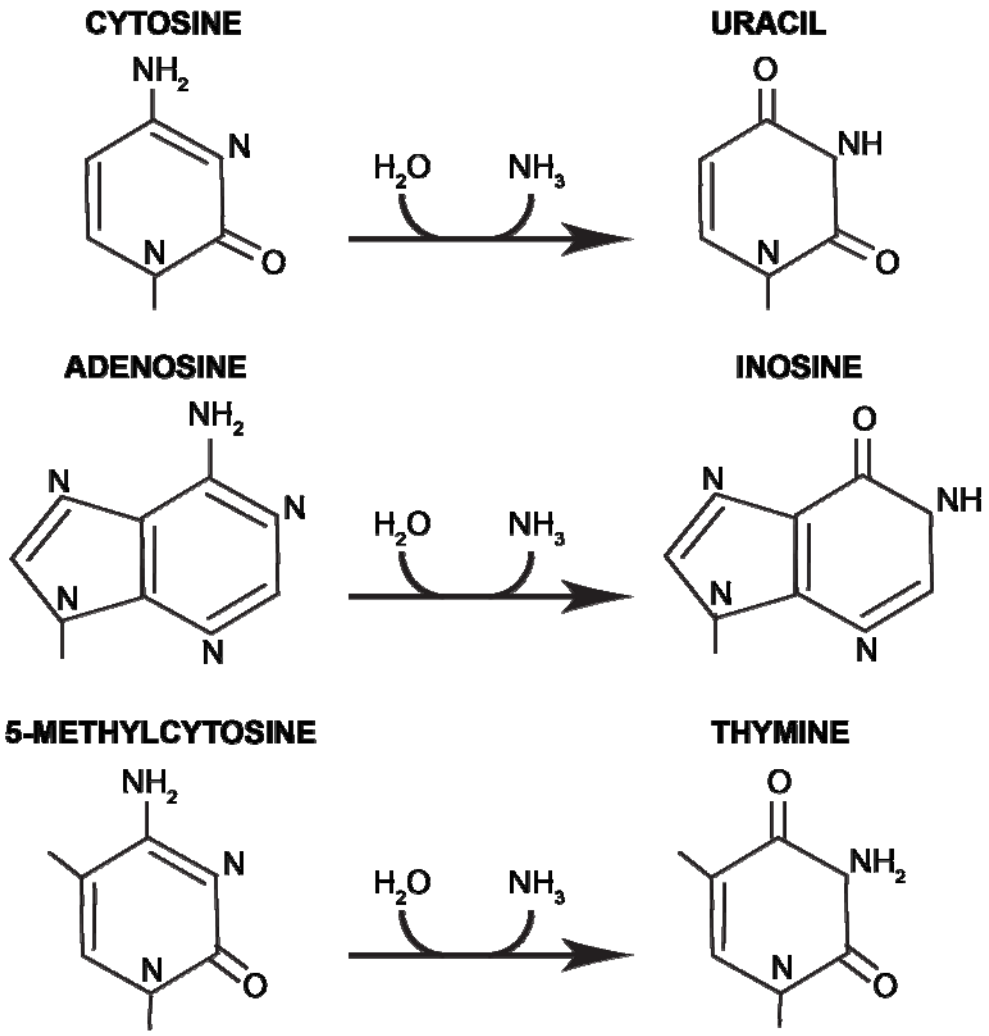


Figure 1.1: Hydrolytic deamination
Hydrolytic deamination of nucleotides by enzymes containing a catalytic cytidine deaminase domain requires water and produces ammonia. Deamination of cytosine produces uracil, of adenosine produced inosine, and of 5-methylcytosine produces thymine.

1.2 Structure and evolution of the cytidine deaminase domain

The crystal structure of the CDA domain from two key enzymes has been solved. The structures of murine adenosine deaminase (ADA) (Wilson *et al.*, 1991) and *E.coli* cytidine deaminase (CDA) (Betts *et al.*, 1994) revealed a zinc atom within the active site and four conserved residues that coordinate the zinc atom. These residues are highly conserved throughout the CDA family and are essential for catalytic activity. ADA and CDA enzymes both deaminate free nucleosides through

hydrophilic attack, although ADA targets carbon 6 of a purine ring and CDA targets carbon 4 of a pyrimidine ring.

Sequence alignment revealed that there was an expanded family of deaminases that contain the conserved CDA domain (Gerber and Keller, 2001). The conserved residues within the CDA domain are: a cysteine (C) or histidine (H) residue followed by a glutamic acid (E) residue which acts as a proton donor during the nucleophilic deamination reaction; a proline (P) followed by two conserved cysteine residues which coordinate the zinc atom in the active site; and a conserved secondary structure of alternating α helices and β sheets. These conserved residues are organised into three motifs: motif I (H/CxE), motif II (PC) and motif III (C), where x indicates any amino acid (see Figure 1.2 for alignment of the CDA domain).

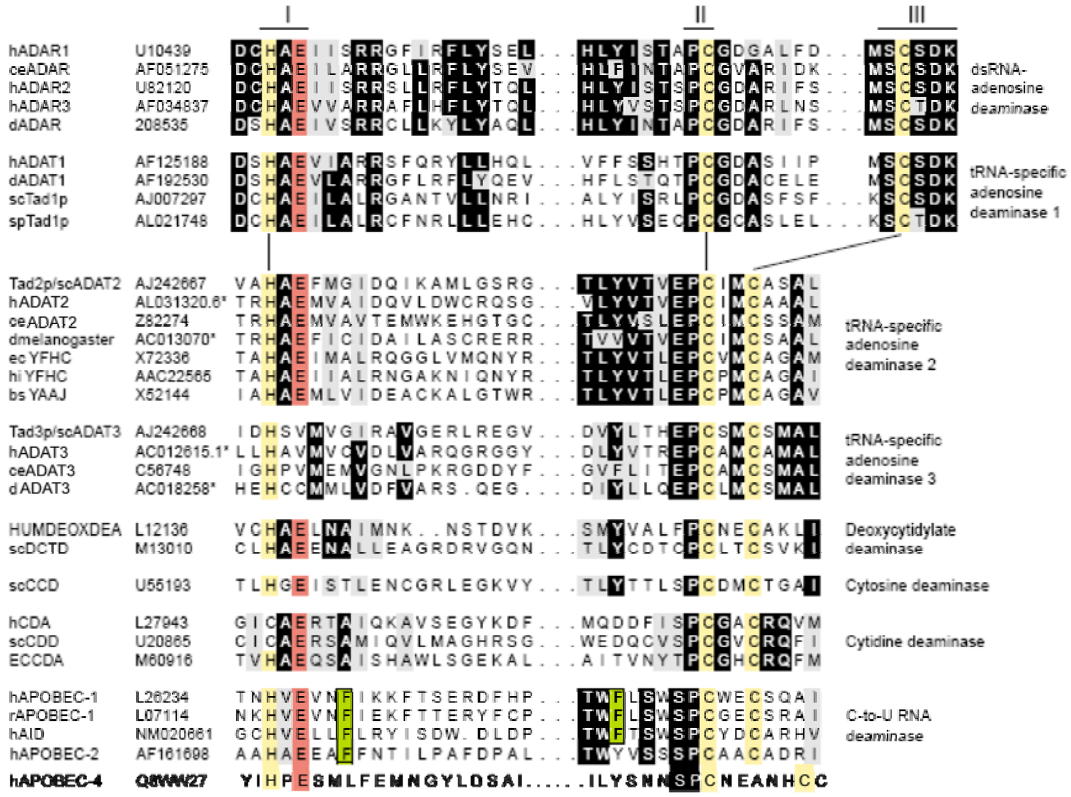


Figure 1.2: Alignment of CDA domains
Protein accession numbers are listed on the left and the deaminase motifs I, II, and III are indicated above. Highly conserved residues shared within a class of enzymes are highlighted in black, semi-conserved ones are grey, residues that coordinate the zinc atom in the active site are yellow, the catalytic site glutamic acid is highlighted in red, and conserved phenylalanine residues in the AID/APOBEC family are green. A4 shares some of the crucial residues conserved across the CDA superfamily but does not contain residues that are conserved within the AID/APOBEC family. Figure modified from Gerber & Keller, 2001.

The CDA domain is shared throughout several diverse families which deaminate adenosines or cytidines within an RNA or DNA substrate (Gerber and Keller, 2001). The conserved domain indicates these processes are linked despite their obvious differences, most notably the different substrates. ADAR family members deaminate adenosine residues within a double-stranded RNA substrate whereas members of the AID/APOBEC family deaminate cytosine residues from single stranded RNA or DNA substrates.

APOBEC2 (A2) was the first member of the AID/APOBEC family to have its crystal structure solved (Prochnow *et al.*, 2007). The structure was solved as a tetramer, although in this formation the active site is not accessible indicating this structure is either unique to A2 or the protein may be active as a dimer (Bransteitter *et al.*, 2009). Furthermore the *E.coli* CDA functions as a dimer (Betts et al., 1994) and APOBEC1 (A1) is believed to form a dimer through the leucine-rich, carboxy-terminal domain (Lau *et al.*, 1994). APOBEC3G (A3G) can form multimers and is believed to function as a tetramer, however it was recently shown by mutational analysis that A3G monomers are catalytically active (Opi *et al.*, 2006). But this could indicate that the model used to infer crucial residues involved in multimerisation for mutational inactivation is incorrect. Some members of the APOBEC3 (A3) sub-family contain two deaminase domains; therefore dimers of these proteins may have the structure and appearance of tetramers of single deaminase domain proteins which complicates the analysis.

ADARs function as dimers (Cho *et al.*, 2003; Gallo *et al.*, 2003; Jaikaran *et al.*, 2002; Macbeth *et al.*, 2004; Poulsen *et al.*, 2006; Valente and Nishikura, 2007), however the occurrence of heterodimers between different family members remains controversial (Chilibeck *et al.*, 2006; Cho et al., 2003). The CDA superfamily deaminates residues within double-stranded oligomers. These can be formed intermolecularly or intramolecularly where the complementary region is termed the editing complementary sequence (ECS) (Higuchi *et al.*, 1993).

1.3 The ADATs

It has long been known that inosine base occurs at the "wobble" position of several tRNA anticodons, however little was known about the process of generating inosine residues in tRNA until discovery of a tRNA-specific adenosine deaminase (Tad1p/ADAT1) in *Saccharomyces cerevisiae* (Gerber *et al.*, 1998). ADAT1 was identified through homology to ADARs, although it lacks the characteristic double-strand RNA binding domains (dsRBDs) of other family members. ADAT1 was shown to deaminate A-to-I at position 37 of $tRNA^{Ala}$, however the Tad1p gene is not essential and ADAT1 is not able to edit the "wobble" position.

Editing of an adenosine residue at the "wobble" position of the anticodon (position 34) allows the third base of the anticodon to recognise adenine, cytosine or uracil in the messenger RNA (mRNA) thus reducing the number of tRNA molecules required for translation. Further investigation showed that this editing event is performed by a heterodimeric enzyme complex comprised of an ADAT2 and an ADAT3 monomer, with the catalytic activity provided by ADAT2 (Gerber and Keller, 1999). Both genes encoding these subunits are essential indicating the importance of inosine at the "wobble" position. ADAT2 contains a conserved CDA domain and is catalytically active, whereas ADAT3 lacks crucial residues from the CDA domain and is not catalytically active. Interestingly, a recent paper describes a tRNA-editing enzyme purified from *Trypanosoma brucei* (TbADAT2) that can deaminate both A-to-I in tRNA and C-to-U in DNA, providing direct evidence these enzymatic activities are related (Rubio *et al.*, 2007). However the TbADAT2 enzyme showed some specificity as it was not able to deaminate cytosine in RNA or adenosine in DNA. Some members of the AID/APOBEC family can deaminate cytidine residues in either DNA or RNA, and evolved from ancestral ADAT enzymes in the vertebrate lineage.

1.4 The AID/APOBEC family

The AID/APOBEC family of cytidine deaminases in mouse includes APOBEC1 (A1), APOBEC2 (A2), APOBEC3 (A3), Activation Induced Deaminase (AID) and the newest putative member APOBEC4 (A4; see Figure 1.3). Expansion of the AID/APOBEC family occurred around the same time as vertebrate divergence and adaptive immunity evolved. Phylogenetic analysis revealed that AID and A2 are the ancestral members of the AID/APOBEC family; only one copy of AID is present but two copies of A2 exist in the Zebrafish genome likely due to genome duplication events. AID and A2 are located in syntenic chromosomal regions across human, mouse and chicken genomes (Conticello *et al.*, 2005). In contrast, A1 is a mammal-specific gene that is likely to have arisen from a duplication of AID, with a distinct C-terminal region that displays the most divergence from AID. It is now proposed that the ability of A1 to edit mRNA may have evolved after it diverged from the ancestral AID/A2.

It is likely that the A3 locus also originated from a duplication of the AID gene, and the region shows synteny between mouse and human despite a massive expansion from one copy in mouse to seven copies in primates (Jarmuz *et al.*, 2002). The seven copies of human A3 (A, B, C, DE, F, G, and H) present at one locus on chromosome 22 have undergone extensive rearrangements and are interspersed with repetitive element sequences. This indicates that retrotransposition of the repetitive elements which included the neighbouring genes may be responsible for the amplification of the A3 genes in the primate lineage (Conticello et al., 2005; Jarmuz et al., 2002). There is also another A3-like gene in humans on 12q24.11, however this gene contains no introns and no expressed sequence tags (ESTs) have been annotated to this gene, indicating it is probably a pseudogene. Mouse A3 contains two deaminase domains indicating a duplication event followed by a fusion of two copies of the gene occurred in the A3 lineage. During the primate-specific expansion four A3 genes evolved containing two deaminase domains (A3B, A3DE, A3F, and A3G) whilst three genes retain only one deaminase domain (A3A, A3C, and A3H). A3 proteins containing two deaminase domains are much larger than those with one deaminase domain which has implications for their subcellular localisation. It has

been demonstrated that the A3 genes are under strong positive selection, which is commonly seen in genes involved in host-pathogen interactions. Surprisingly the positive selection of A3G pre-dates evidence of lentiviral infection of the genomes studied, indicating the A3 genes are evolving in response to other pathogens (Sawyer *et al.*, 2004).

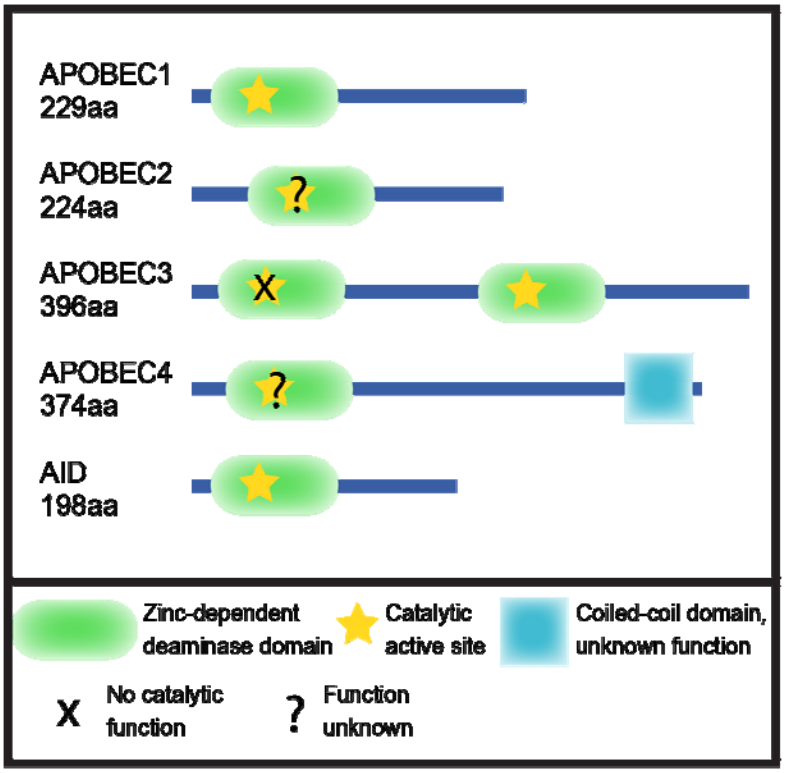


Figure 1.3: Cartoon depicting the members of the mouse APOBEC/AID family.
All five members of the APOBEC/AID family share a conserved deaminase domain (green) containing the catalytic active site (yellow star), however the active site is not catalytically active. In the case of A3 which contains 2 deaminase domains only the one in the carboxy region has catalytic activity. A2 has no known substrates and does not display catalytic activity when assayed *in vitro* however this could be due to the conditions used. The catalytic activity of APOBEC4 has yet to be determined. A4 is the only member of the family to contain a short coiled-coil domain in the carboxy-terminal region, its biological function is unknown.

APOBEC1

APOBEC enzymes were first discovered during the study of Apolipoprotein (Apo) B, a large glycoprotein which has important roles in lipid metabolism and the regulation of lipoprotein levels in plasma (Powell *et al.*, 1987). Two active isoforms of the ApoB protein occur in mammals: one corresponds to the full length transcript

(ApoB100); the other is encoded by the 2,152 amino terminal residues (ApoB48). ApoB48 is generated via a single C-to-U editing event in the ApoB mRNA at cytosine nucleotide number 6666, which converts a glutamine codon (CAA) to a premature stop codon (UAA). This editing reaction is catalysed by APOBEC1 (A1), a 27kDa protein which is a member of the AID/APOBEC group of the CDA family (Navaratnam *et al.*, 1993; Teng *et al.*, 1993). ApoB editing is functionally important due to its role in atherosclerosis. Lipoprotein particles containing ApoB100 are required for the production of very low density lipoprotein (VLDL) particles, which are then metabolized to intermediate density lipoprotein (IDL) and low density lipoprotein (LDL). Whereas, lipoprotein particles containing ApoB48 are metabolised faster and are not processed into LDL particles. LDL particles transport plasma cholesterol around the body, and LDL levels have been shown to be one of the major risk factors for coronary heart disease. The role of ApoB in atherogenesis and the process of RNA editing are reviewed in (Davidson and Shelness, 2000).

A1 is expressed in the small intestine in mammals where editing of ApoB mRNA occurs to 70-95%, however in rodents editing of ApoB also occurs in the liver to approximately 60%. Two different transcripts are produced from the rat *Apobec1* locus due to alternative splicing and different promoter use; one isoform of 1kb expressed in the small intestine and one of 1.24kb expressed in the liver (Teng et al., 1993).

Apobec-1 knockout mice lack editing of the *ApoB* transcript and no ApoB48 protein is detected in plasma, but the mice are viable and fertile (Hirano *et al.*, 1996). A more detailed examination of *Apobec1*-null mice revealed that the absence of ApoB48 was accompanied by an increase in the amount of ApoB100 protein, which produced an imbalance in the proportion of high density lipoprotein (HDL) to LDL (Morrison *et al.*, 1996; Nakamuta *et al.*, 1996). This mild phenotype was unexpected, and indicates the primary function of A1 is in modulating lipid metabolism through editing of the ApoB mRNA, and this function cannot be performed by another enzyme in the absence of A1.

To determine whether editing of the ApoB transcript is developmentally regulated mice were engineered to endogenously express either ApoB48 or ApoB100 (uneditable). However, these mice developed normally and were healthy and viable, although differences in plasma LDL levels remained (Farese *et al.*, 1996), indicating the editing of ApoB mRNA is dispensable during mouse development. Current thinking supports the hypothesis that editing of the ApoB mRNA is regulated in response to changing hormone levels via an increase in *Apobec1* expression. Indeed, *Apobec1* expression is increased, as is editing of ApoB mRNA, in rat primary hepatocytes in response to insulin treatment (Sowden *et al.*, 2002).

Yamanaka *et al*, noticed a correlation between animals that edit ApoB transcripts in the liver and lower levels of plasma LDL and reasoned that expression of A1 in the liver may be sufficient to reduce LDL levels (Yamanaka *et al.*, 1995). Animals engineered to express *Apobec-1* in the liver developed liver dysplasia and tumours similar in profile to hepatic carcinomas. Non-specific editing of mRNAs not known to be substrates for A1 was demonstrated and hypothesised to be the cause of the tumour development (Yamanaka et al., 1995). As yet no other *in vivo* substrates have been found for A1, but this study reveals that it is capable of editing endogenous liver mRNAs.

Editing of ApoB mRNA occurs in the nucleus immediately after transcription, and is probably coupled with the processes of splicing and polyadenylation (Lau *et al.*, 1991). Editing of the ApoB mRNA by A1 requires the correct secondary structure of the RNA sequence surrounding the edited position. Mutation analysis revealed a critical "mooring sequence" of 11 nucleotides which occur 3' of the edited cytosine and form a stem-loop structure (Backus and Smith, 1991). An upstream enhancer region was also identified and the spacing between these *cis* elements was found to be important for efficient editing (Backus and Smith, 1992).

Editing of ApoB mRNA requires both the catalytic A1 deaminase and at least one co-factor called APOBEC1 complementation factor (ACF) (Mehta *et al.*, 2000). The ACF transcript is alternatively spliced to produce two isoforms of 64 and 65kDa in

mice, which differ in the inclusion of one exon encoding 8 amino acids. Interestingly ACF expression is widespread, including many tissues that lack ApoB and *Apobec1* expression suggesting it has other biological functions (Mehta et al., 2000). ACF is present in both the cytoplasm and associated with heterochromatin in the nucleus. ACF is a member of the hnRNP family and contains three RNA recognition motifs (RRMs), through which it interacts with AU-rich sequences in the ApoB mRNA and also with A1. ACF contains a NLS and can shuttle in and out of the nucleus via interaction with transportin 2, indeed ACF is required for the localisation of A1 to the nucleus for editing of ApoB mRNA (Blanc *et al.*, 2001). However, *in vitro* studies have shown that A1 is capable of editing ApoB mRNA without its cofactor ACF (Chester *et al.*, 2004), indicating the role of ACF *in vivo* may be to localise A1 to the nucleus.

ACF null mice are embryonic lethal, with embryos unable to survive past the blastocyst stage (E3.5) indicating that ACF has crucial functions distinct from its role in the editing of ApoB mRNA (Blanc *et al.*, 2005). Heterozygous mice are viable, fertile and healthy despite lower levels of ACF protein, and surprisingly they displayed a higher level of ApoB RNA editing in the liver and kidney though levels in the small intestine were unchanged. Cell culture experiments indicated loss of ACF resulted in an increase in apoptotic cell death, and this effect was independent of *Apobec-1* expression (Blanc et al., 2005). Taken together these results indicate ACF has a crucial role early in development which is independent of ApoB editing.

To further investigate the DNA deaminase activity of A1 and other members of the AID/APOBEC family, an *E.coli* DNA mutator assay was developed (Harris *et al.*, 2002; Petersen-Mahrt *et al.*, 2002). The assay relies on mutation of a gene (in this instance *rpoB*) giving rise to a positive selectable marker (rifampicin resistance, the assay is described in detail in (Coker *et al.*, 2006)). The assay was used successfully to demonstrate that A1, as well as AID, A3C and A3G can deaminate cytosine residues in DNA, whilst in this assay A2 could not deaminate DNA (Harris et al., 2002). Mutation rates were significantly increased in uracil DNA glycosylase-deficient *E.coli* (*Ung-1*); indicating the resistance was due to deamination of cytidine

to uracil. Indeed, sequence analysis showed the enzymes displayed distinct preferences for the sequence context of the cytidine to be deaminated. Recent work on A1 has shown it is capable of deaminating ssDNA *in vitro*, indicating it may have a more varied biological role than previously described (Petersen-Mahrt and Neuberger, 2003). Interestingly A1 and AID can also deaminate 5-methylcytosine to thmidine in a modified *E.coli* DNA mutator assay, where a CpG-directed methyltransferase is co-expressed in a methylation restrictive defective *E.coli* strain (Morgan *et al.*, 2004).

APOBEC2

A2 is an ancestral member of the AID/APOBEC family, thought to have originated from a duplication of the AID gene in bony fish (Conticello et al., 2005). Expression of A2 is restricted to heart and skeletal muscle, however it was unable to substitute for A1 in an ApoB mRNA editing assay and little is known about its function or substrates *in vivo* (Liao *et al.*, 1999). Biochemical investigation showed that recombinant A2 has little or no affinity for either non-specific AU rich RNA or Apo-B RNA *in vitro*, indicating it does not function as an RNA deaminase. Deletion of the gene in mice produced apparently healthy, fertile mice with no immediately obvious phenotype (Mikl *et al.*, 2005). Further investigation of the A2-null mice revealed a muscle-specific phenotype that increased in severity with age resulting in a mild myopathy (Sato *et al.*, 2009). This was accompanied by a 15-20% decrease in total body mass, evident from birth. The role of A2 remains unclear however a clear myopathic phenotype albeit mild, in the A2 knockout mice indicates a function in skeletal muscle. This may be mediated through nucleic acid binding rather than cytidine deaminase activity, and substrates remain to be identified.

APOBEC3

The A3 subfamily consists of seven proteins in primates (APOBEC3- A, B, C, DE, F, G, and H), which are potent inhibitors of retroviruses. Several A3 genes have

been shown to be expressed in human tumour cell lines when expression is not normally detected in the tissues they are derived from (Jarmuz et al., 2002).

Anti-retroviral activity was initially observed when HIV-1 mutants were isolated that lacked the virion infectivity factor (Vif). These mutants failed to replicate in certain CEM (T-cell) cell lines (termed "non-permissive") but replicated normally in other CEM cell lines (termed "permissive"). Sheehy *et al* (2002) used a cDNA library subtraction screen to identify a host factor APOBEC3G/CEM15 (A3G) as the protein that conferred the non-permissive phenotype when expressed in permissive cells (Sheehy *et al.*, 2002). Initial investigations demonstrated that the inhibition occurs via A3G interacting with the viral gag protein which is a component of the viral nucleocapsid. A3G is subsequently incorporated into the new virus particle along with the viral RNA (Alce and Popik, 2004; Zennou *et al.*, 2004). Following infection of a new cell viral RNA is reverse transcribed into DNA and A3G deaminates cytidine residues in the single-stranded minus strand of the nascent proviral DNA (Yu *et al.*, 2004). The resulting uracil residues are excised by virion-specific uracil DNA glycosylase-2 (UNG2) and the proviral DNA is degraded by apyrimidic endonuclease (APE) (Yang *et al.*, 2007). Therefore A3G is a potent inhibitor of HIV-1 in the absence of Vif, and as such A3G provides a powerful potential therapeutic in the fight against HIV-1 and AIDS.

However, HIV-1 counteracts the anti-retroviral activity of A3G through the virally encoded protein Vif. Further studies to characterise this interaction revealed the HIV-1 Vif protein can counteract the antiviral effects of A3G in several ways: Vif binds and sequesters A3G preventing it from being packaged into virions (Mariani *et al.*, 2003); Vif promotes A3G ubiquitination which targets it for proteasome-mediated degradation (Kobayashi *et al.*, 2005); and Vif directly inhibits translation of A3G mRNA (Kao *et al.*, 2003). In combination these measures effectively ensure that no A3G is incorporated into budding virions. The Vif-A3G interaction is highly species specific, as it has been shown that HIV-1 Vif protein could not inhibit mouse A3 (Mariani et al., 2003). Further investigation revealed that HIV-1 Vif could not inhibit A3G from species that are evolutionarily much closer to humans such as

African green monkeys and macaques (Mariani et al., 2003). Conversely, human A3G is not susceptible to inhibition by simian immunodeficiency virus (SIV) Vif protein. Several groups managed to refine the species-specific inhibition by HIV-1 Vif to one critical residue, aspartic acid 128, which corresponds to a lysine residue in African green monkey A3G (Bogerd *et al.*, 2004; Mangeat *et al.*, 2004; Schrofelbauer *et al.*, 2004; Xu *et al.*, 2004). Mutating human A3G amino acid 128 from aspartic acid to lysine can render it susceptible to SIV Vif but resistant to HIV-1 Vif. This level of specificity indicates that A3G has undergone recent evolution, possibly in response to rapid evolution of retroviruses. A3G also displays antiviral activity against Hepatitis B virus (HBV), which undergoes a reverse transcription step during replication (Turelli *et al.*, 2004). However, in this case A3G interferes with packaging of the viral pre-genome, and there is no evidence of hypermutation.

The antiviral activity of A3 is not specific to the primate lineage. A mouse locus outside of the major histocompatibility complex involved in recovery following infection with Friend Virus (FV) termed *Recovery from Friend Virus 3* (*Rvf3*) was recently found to encode mouse A3 (Santiago *et al.*, 2008). The *Rvf3* viral resistance factor locus was identified as an autosomal dominant trait in 1979, which resulted in reduced viral load and increased neutralising antibodies following infection with FV (Chesebro and Wehrly, 1979). Susceptible mouse strain BALB/c was found to produce an alternatively spliced A3 transcript lacking exon 2, which displayed reduced antiviral activity in *in vitro* studies. Therefore mouse A3 contributes to reduced viral load *in vivo*, but also appears to stimulate neutralising antibodies indicating a wider role for A3 in the immune system. The authors concluded that the absence of the ability to stimulate an antibody reaction following HIV-1 infection may be due to the presence of Vif. Mouse A3 is predominantly expressed in lymphoid tissues and knockout mice display no obvious phenotype (Mikl et al., 2005). However further investigation revealed that mouse A3 inhibits the murine Mammary Tumour virus (MMTV) replication *in vivo*, such that A3-null mice were more susceptible to MMTV infection (Okeoma *et al.*, 2007).

APOBEC3 and transposable elements

Only a small fraction of cellular A3G is incorporated into the budding HIV-1 virions and this prompted investigation into alternative cytoplasmic roles for A3G. Chiu and colleagues observed that activation of CD4+ T cells was accompanied by a shift in A3G from a low molecular mass complex to a high molecular mass complex, and this shift coincided with a loss of antiretroviral activity (Chiu *et al.*, 2005). The shift could be reversed by treatment with RNase A, which caused the high molecular weight complex to dissociate. Analysis of the high molecular weight complexes revealed that they were Staufen-containing RNA transport granules, Ro ribonucleoprotein (RNP) complexes and pre-spliceosomes. Analysis of the RNA species present in the RNP complexes revealed that they contained Alu and small Y nonautonomous retroelement RNAs, as well as mRNAs encoding both A3G and A3F amongst others. Further experiments showed that activation of T cells induces expression of retroelements. The retroelement RNA intermediates are bound by A3G and sequestered into high molecular mass RNP complexes. This process is independent of cytidine deaminase activity, and results in a loss of antiretroviral activity against the HIV-1 virus as the total cellular A3G is now associated with the high molecular mass RNP complexes (Chiu *et al.*, 2006). These findings demonstrate A3G is able to inhibit endogenous retrotransposons through binding and sequestering their RNA intermediates. Therefore A3G has a role as an RNA-binding protein although deamination of RNA was not observed. Further, these results indicate a potential mechanism of regulating A3G activity through its association with high molecular mass RNP complexes. Localization studies indicated these high molecular weight RNP complexes are cytoplasmic RNA processing (P) bodies, where translationally silenced mRNAs can be stored or degraded (Wichroski *et al.*, 2006). A3G could be relocated to stress granules following heat shock, along with specific mRNAs and components of the RNA-induced silencing complex (RISC) (Gallois-Montbrun *et al.*, 2007; Kozak *et al.*, 2006). These results indicate A3G may have additional biological roles through its interaction with cellular mRNAs in P bodies.

The discovery that A3 proteins are capable of restricting retroviruses and inhibiting Alu elements retrotransposition led to investigation of the role of APOBEC proteins

in the inhibition of other endogenous retrotransposons. Retrotransposon-derived sequences account for approximately 45% of the human genome and the three major classes that are still active replicate through an RNA intermediate (Lander *et al.*, 2001). Long terminal repeat (LTR) retrotransposons contain viral *gag* and *pol* genes that are flanked by repeat sequences which control transcriptional regulation and are capable of autonomous retrotransposition. Retroviruses, like HIV-1, are thought to have originated from LTR-retrotransposons by the acquisition of the *env* gene which encodes the envelope proteins, and they replicate in a similar manner to retroviruses utilising a reverse transcriptase to create an RNA intermediate. Long interspersed elements (LINEs) can also replicate autonomously and contain an RNA polymerase II promoter upstream of two open reading frames (ORFs). Following translation the LINE proteins bind to its RNA and the LINE RNA is reverse transcribed in a process termed "target-site primed reverse transcription", which often fails to complete resulting in many 5' truncated inactive repeat elements. LINEs comprise approximately 21% of the human genome, the most common being the LINE-1 (L1) element (Lander et al., 2001). Short interspersed elements (SINEs) are similar to LINEs but cannot replicate autonomously, however LINE encoded proteins can function in *trans* to replicate SINE elements. SINEs account for approximately 13% of the human genome, yet only one family remains active which is also the most abundant; the *Alu* element (Lander et al., 2001).

Retrotransposition events can have serious consequences for the host genome, not least through the accumulation of genetic material. Several potential consequences are described here, for a comprehensive review see (Cordaux and Batzer, 2009). Consequences of retrotransposition include insertional mutagenesis, whereby a retrotransposon can insert into the coding sequence or regulatory elements of a gene, or insertion-mediated deletion where genetic information is lost from regions adjacent to the insertion site. Retrotransposons can also induce genomic rearrangements whereby recombination events take place between homologous retrotransposon sequences on different chromosomes. During replication retrotransposons can copy flanking genes, or parts thereof, and integrate them locally or elsewhere in the genome (this occurred during the primate specific expansion of

A3 genes which is described earlier). Whilst sometimes these rearrangements can be beneficial, the potential for deleterious mutation is great, therefore host genomes have evolved mechanisms to limit retrotransposition.

There are numerous studies on the effect of APOBEC proteins on retrotransposition, some of which are described below, others are summarised in Table 1.1. A3G and A3F were shown to inhibit retrotransposition if the LTR-retrotransposon Ty1 in *Saccharomyces cerevisiae*, through deamination of the reverse transcribed DNA intermediate (Schumacher *et al.*, 2005). Investigation of mouse LTR-retrotransposons Intracisternal A-particles (IAPs) revealed mouse A3 and human A3G were able to significantly inhibit retrotransposition via hypermutation of the DNA intermediate, however mouse A1 and A2 had no effect (Esnault *et al.*, 2005). Another study into inhibition of IAPs showed that A3G and A3B reduced retrotransposition by 4-fold and 20-fold respectively (Bogerd *et al.*, 2006a). This inhibition was mediated by association of A3G or A3B with the IAP Gag protein which resulted in the inclusion of the A3 protein into the virus-like particle where it hypermutated the nascent cDNA. However, A3A was shown to reduce retrotransposition of the murine IAP by 100-fold yet failed to associate with the IAP Gag protein *in vivo*, and IAP transcripts analysed did not show hypermutation indicating that A3A inhibits retrotransposons by a novel cytidine deamination-independent mechanism (Bogerd et al., 2006a).

These studies show that A3s are effective against LTR-containing retrotransposons, however initial investigations into inhibition of L1 and *Alu* retrotransposition indicated that only A3A and A3B were capable of inhibiting these elements through a deamination-independent mechanism (Bogerd *et al.*, 2006b). Further studies including all human A3s (A-H) showed that they are all capable of inhibiting L1 retrotransposition, although some are more effective than others (Kinomoto *et al.*, 2007). Thus the same mechanism of deaminating the nascent cDNA strand resulting in hypermutation and degradation of proviral DNA is observed in the inhibition of both retroviruses and retrotransposons. However, additional cytidine deamination-independent mechanisms are also involved in the inhibition of retrotransposons.

Investigation into the mechanism of A3B and A3F inhibition of L1 retrotransposition revealed key catalytic site residues were dispensable and no evidence of cytidine deamination was found (Stenglein and Harris, 2006). Inhibition of L1 retrotransposition by A3B and A3F resulted in fewer copies of integrated L1 DNA indicating inhibition occurred before integration; however the precise mechanism of inhibition is as yet undetermined. This data and other similar results have led to an increase in investigations into the function of A3 proteins and how they are able to inhibit different viruses and retroviruses. Thus, the expanded A3 family present in the primate lineage has an important role in the immune system protecting the genome from endogenous retrotransposons and exogenous retroviruses.

Recent data indicates that A3 proteins may function at several different stages of the innate immune response to infection. The discovery that A3A has limited antiviral activity against HIV-1, but is capable of restricting retrotransposition of *Alu* elements prompted further research into its biological function. A3A is expressed in macrophages and can be induced by both interferon and foreign DNA leading Stenglein *et al* to hypothesise that it is involved in the immune system response to foreign DNA (Stenglein *et al.*, 2010). They showed that A3A can extensively deaminate foreign dsDNA with up to 97% of cytidines being converted to uracil, resulting in degradation of the foreign DNA through UDG-mediated cleavage. This was demonstrated with a striking experiment showing the transfection of a plasmid encoding A3A leads to progressive loss of expression of A3A as it targets its own plasmid for degradation. This makes further studies difficult and past ones open to re-interpretation. Furthermore, despite the very high level of editing detected there was no evidence of endogenous DNA deamination, only foreign DNA was targeted for deamination by A3A indicating a stringent mechanism for recognition of foreign DNA. This study highlights the diverse roles APOBEC proteins play in the immune system, and suggest further investigations are necessary to elucidate the specific biological substrates of each family member.

	HIV-1	MLV	HBV	IAP	Alu	L-1	dsDNA	ssDNA	5MeC	RNA	Subcell. Local[n]
APO1			✓	×		×		✓	✓	✓	nuc/cyto
APO2			×	×		×		×			
APO3A	×	×		✓	✓	✓	✓				nuc/cyto
APO3B	✓	✓		✓	✓	✓	✓				nuclear
APO3C	weak	✓	✓		✓	weak	weak	✓			nuc/cyto
APO3DE	weak	×				weak	✓				cyto
APO3F	✓	×	✓			✓	✓				cyto
APO3G	✓	✓	✓	✓	✓	✓	×	✓			cyto
APO3H	×	×	✓			weak	×				nuc/cyto
AID						✓		✓	✓		nuc/cyto
Refs	summarised in (Kinomoto et al., 2007)	(Doehle *et al.*, 2005; Kinomoto et al., 2007)	(Gonzalez *et al.*, 2009; Kock and Blum, 2008; Suspene *et al.*, 2005)	(Bogerd et al., 2006a; Esnault et al., 2005)	(Bogerd et al., 2006b; Chiu et al., 2006)	(Bogerd et al., 2006b; Esnault et al., 2005; Kinomoto et al., 2007; MacDuff *et al.*, 2009)	(Stenglein *et al.*, 2010)	(Harris et al., 2002; Petersen-Mahrt et al., 2002)	(Morgan et al., 2004)	(Navaratnam et al., 1993; Teng et al., 1993)	summarised in (Kinomoto et al., 2007)

Table 1.1: Summary of activity of APOBEC proteins

Activity of APOBEC proteins tested is listed where ✓ indicates positive, × indicates negative and blank indicates not tested. References are listed, some references describe later studies including comparisons of several family members are written as "summarised in".

Methylation of transposable elements

The consequences of retrotransposition events are rarely beneficial to the host organism and germline insertions play a role in heritable diseases including haemophilias and Duchenne Muscular Dystrophy (Chen *et al.*, 2005). Therefore host organisms have evolved complex mechanisms for preventing retrotransposition events which include silencing transcription through modification of DNA sequences, reviewed in (Maksakova *et al.*, 2008). These modifications consist of methylation of CpG dinucleotides within the promoter region of the retrotransposons and remodelling of the core histones to make the DNA less accessible to transcription factors. Methyl groups are attached to cytosine residues by DNA methyltransferase enzymes (Dnmts) and this causes transcriptional repression through binding of proteins containing methyl binding domains (MBDs). MBD proteins can recruit chromatin modifying enzymes such as histone methyltransferases and histone deacetylases, which modify the N-terminal tails of core histone proteins leading to a more compact chromatin structure. These modifications act in concert to silence transcription through establishing and maintaining an inactive chromatin state. Indeed, artificial targeting of Methyl CpG-binding Protein 2 (MeCP2) to the L1 promoter region effectively represses L1 expression and retrotransposition (Yu *et al.*, 2001).

Activation Induced Deaminase

Activation Induced Deaminase (AID) has been shown to have a crucial role in the adaptive immune response where it is required for several distinct steps in the affinity maturation of antibodies during an immune response (summarised in Figure 1.4). The first step of antibody maturation occurs without AID, and involves homologous recombination to select one segment from each of three genes (variable V, diversity D, and joining J genes). V(D)J recombination occurs between recombination signal sequences located either side of the gene segments and results in a wide repertoire of low affinity antibodies that are all of class IgM, the process is reviewed in (Maizels, 2005).

AID functions in the next step of antibody affinity maturation, which is class switch recombination (CSR, Figure 1.3 middle panel). This results in an intrachromosomal deletion of the IgM constant region and its replacement by a downstream constant region gene producing antibodies with different constant regions. CSR is dependent on AID, which is demonstrated in patients deficient in AID that display hyper-IgM-2 syndrome, who only produce IgM antibodies as they are unable to undergo CSR (Revy *et al.*, 2000). AID is believed to initiate double strand breaks (DSBs) in DNA required for CSR via deaminating cytosine residues in the immunoglobulin S region to uracil residues. The uracil is recognised by a uracil DNA glycosylase enzyme (for example UNG) and subsequently removed, either by UNG or by mismatch repair (MMR) enzyme MSH2, leaving an abasic site exposed (Rada *et al.*, 2004). Repair of staggered DSBs occurs through non-homologous end-joining (NHEJ) and produces functional antibodies of different classes from one locus.

AID is also involved in the process of somatic hypermutation (SHM) which allows the immune system to produce a wide repertoire of antibodies with different effector function following activation of naïve B cells by antigen. This process of affinity maturation is achieved through inducing random point mutations within a 1.5-2kb region spanning the recombined V(D)J section, which encodes the antigen-binding region of the antibody (Lebecque and Gearhart, 1990). The mechanism whereby AID is targeted to the immunoglobulin locus remains elusive, although histone modifications are thought to create an accessible chromatin environment (Nambu *et al.*, 2003) and *cis* acting sequence elements likely play a role (Zarrin *et al.*, 2005). The mutation rate in this region is in the order of 10^4/base pair/generation, which is considerably higher than any other loci (Berek and Milstein, 1988). The mechanism is similar to that of CSR, however DSBs are not formed. AID deaminates cytidines in the DNA to uracil, which can lead to several different outcomes. If replication occurs before the uracil is detected a transition mutation (dC-dT) occurs. If the uracil is detected by UNG it can be excised leaving an abasic site and replication at this stage across the abasic site results in any base being incorporated in this position. Finally, the mismatched uracil can be repaired through either the base excision repair (BER) pathway or the mismatch repair (MMR) pathway, which can utilise error-

prone polymerases capable of inducing multiple mutations in the surrounding bases. The process of SHM generates a large pool of diverse antibodies, which are then subjected to selective pressure to generate antibodies with higher affinity.

The processes of CSR and SHM described above depend on AID deaminating cytosine residues in DNA within the immunoglobulin locus. This hypothesis is supported by *in vitro* experiments showing that purified AID can deaminate cytosine in single-stranded DNA (Bransteitter *et al.*, 2003). The spectrum of mutations produced *in vitro* demonstrated that AID preferentially deaminates cytidine residues in the motif "WRC" (where W is adenosine or thymine, R is a purine, and C is cytidine), which concurs with the SHM mutation "hotspots" observed *in vivo* (Pham *et al.*, 2003). AID is also capable of mutating DNA when it is expressed in *E.coli* (Petersen-Mahrt et al., 2002) and yeast (Poltoratsky *et al.*, 2004). Yet this is not conclusive evidence as over expression of A1 in *E.coli* results in deamination of DNA when its *in vivo* target is RNA (Harris et al., 2002). Furthermore, SHM requires active transcription from the immunoglobulin locus and AID is localised to the immunoglobulin switch loci during CSR and can be co-immunoprecipitated with RNA Polymerase II (Nambu *et al.*, 2003).

However, there are two potential mechanisms of AID action: AID may act directly on DNA as described above, known as the "DNA-editing hypothesis"; or potentially AID may edit one or more transcripts which encode endonucleases or recombinases that act on the DNA, known as the "RNA-editing hypothesis". To date it has not been conclusively shown which nucleic acid is the substrate of AID.

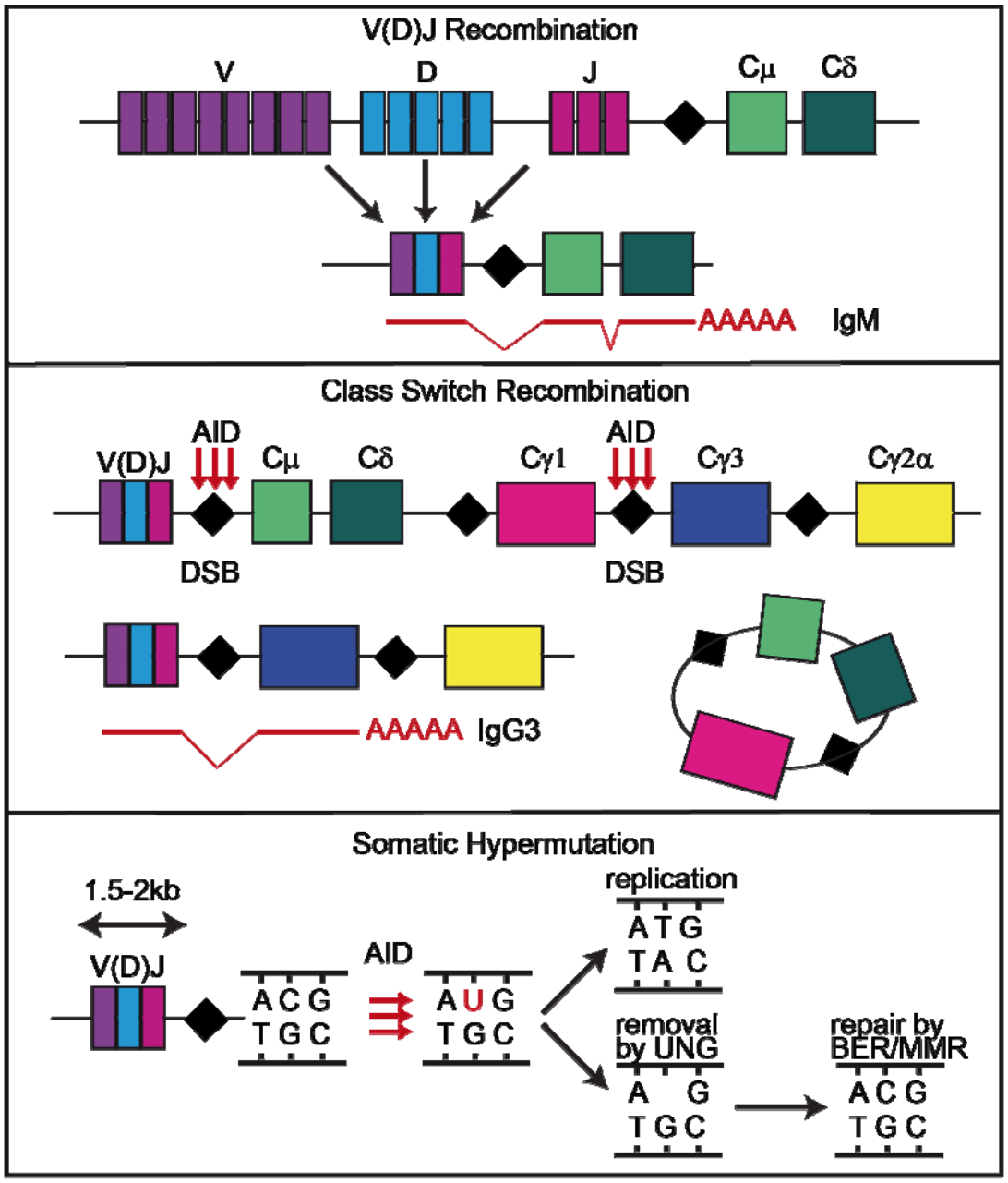


Figure 1.4: Three steps of antibody diversification

Top panel: V(D)J recombination. One segment is selected from each gene, variable (V), diversity (D), and joining (J) by recombination events between recombination signal sequences located either side of each segment. Transcription of this gene generates an IgM class antibody containing the μδ contstant region. This process is not dependent on AID. **Middle panel: Class Switch Recombination.** Deamination of homologous switch regions (black diamonds) located between different constant regions induces double strand breaks (DSBs) which can be repaired by non-homologous end-joining (NHEJ) machinery resulting in production of a different class of antibody (here IgG3), and loss of the DNA section in between. **Lower panel: Somatic Hypermutation.** AID deaminates cytidine residues in a 1.5-2kb region spanning the V(D)J region. The resulting uracil (U) can either be replicated to induce dC-dT mutations, or removed by a uracil DNA Glycosylase enzyme (here UNG) and then repaired by enzymes involved in Base Excision repair (BER) or Mismatch repair (MMR).

Evidence in support of an RNA intermediate comes from experiments showing that *de novo* protein synthesis is required to form double strand breaks in DNA during CSR, following activation of AID (Begum *et al.*, 2004). However, this does not exclude the possibility that AID requires co-factors that must be newly synthesised; therefore current thinking favours the model whereby AID acts directly on DNA at the immunoglobulin locus. Other evidence in support of the RNA-editing hypothesis comes from data which shows AID can be immunoprecipitated in a complex with polyadenylated mRNA *in vivo*, although a direct interaction between AID and RNA was not shown (Nonaka *et al.*, 2009). The authors hypothesise that AID could have an RNA-binding co-factor analogous to A1 and its co-factor ACF; however the interaction of AID with RNA polymerase II (Nambu et al., 2003) could be responsible for the indirect association observed between AID and mRNA. In conclusion, AID is required for efficient antibody affinity maturation likely through deamination of DNA at the immunoglobulin locus, although the mechanism of targeting AID to this locus and the co-factors involved are yet to be determined.

Given AID's role in generating mutation through deamination it was hardly surprising when it was found to be involved in the development and progression of cancer. Reciprocal translocations involving the immunoglobulin locus and a selection of different proto-oncogenes are commonly found in B cell malignancies, such as the *c-Myc-Igh* translocation in Burkitt's lymphoma, reviewed in (Ramiro *et al.*, 2006). The DSBs at the *Igh* locus are characteristic of AID deamination, and AID may also be involved in generating DSBs within the c-myc locus (Robbiani *et al.*, 2009). This results in massive upregulation of the proto-oncogene when the naïve B-cell is stimulated to generate antibodies, and can lead to cancer when tumour-suppressor genes such as p53 are lost.

Similar to recent findings which demonstrated that A3s are capable of inhibiting retrotransposition, AID was shown to inhibit both mouse retrotransposons L1 and MusD, but not the yeast retroelement Ty1 (MacDuff *et al.*, 2009). Intriguingly, the catalytic site glutamic acid residue, and several of the zinc-coordinating cysteine residues were not required for this inhibition. Indeed no evidence of deamination

was detected confirming that the inhibition of LINEs by AID occurs via a deamination-independent mechanism. The authors detected expression, albeit weak, of *Aid* in the ovary indicating inhibition of retroelements in the germline may be another function of AID.

Regulation of AID expression

The genomic region surrounding miR-155 has long been of interest since it was identified as the integration site for proviral DNA from the Avian Leukosis Virus (ALV) in chickens. Expression of a noncoding transcript from this region denoted as *Bic* (B cell integration cluster) was stimulated by the ALV promoter and resulted in oncogenesis (Tam *et al.*, 1997). *Bic* is the primary miRNA precursor from which miR-155 is processed, and subsequently many B cell lymphomas have been found to overexpress either miR-155 or *Bic* (Eis et al., 2005). Indeed transgenic overexpression of miR-155 from an immunoglobulin promoter results in malignant lymphomas (Costinean *et al.*, 2006), whereas mice lacking miR-155 have immune response defects and are unable to produce a full repertoire of antibodies (Thai *et al.*, 2007). Expression of miR-155 is induced following stimulation of B cells to undergo CSR, indicating it plays a role in this process, yet miR-155 is widely expressed in the lymphatic system and is predicted to have over 60 targets *in vivo*. One target of miR-155 is the *Aid* transcript and mice engineered to lack the AID 3'UTR target site for miR-155 produce higher levels of AID, which results in a higher frequency of CSR (Teng *et al.*, 2008). Mice lacking the miR-155 target site in the *Aid* 3'UTR showed expression of *Aid* in a subset of mature circulating B cells where it would normally not be found following stimulation of an immune response. However, this was not a general mis-regulation of *Aid* as peripheral T lymphocytes had correctly switched off *Aid* expression. This indicates AID expression is tightly regulated in both a temporal and a cell-specific manner, and that additional mechanisms exist for regulating *Aid* expression. Despite higher levels of *Aid* these mice did not have increased SHM of immunoglobulin genes but elevated AID in the germinal centre resulted in off-target deaminase activity which induced hypermutation at the oncogene *Bcl6* locus (Teng et al., 2008).

A parallel study revealed the incidence of Burkitt's lymphoma-associated *Myc-Igh* translocations was 3-6 fold higher in mice lacking the miR-155 target sequence (Dorsett *et al.*, 2008). Curiously these mice did not show elevated levels of B cell lymphoma despite harbouring higher levels of the causative translocation, indicating further mutations are required for lymphoma development. Mice ubiquitously expressing AID develop T cell tumours however they do not develop lymphomas which led the authors to postulate that AID activity is also regulated post-transcriptionally in B cells (Muto *et al.*, 2006).

AID as a demethylase

Recent work has shown that AID and A1 can deaminate methylated cytosine residues in an artificial *E.coli* assay resulting in thymine incorporation and a loss of methylation (Morgan et al., 2004). The authors demonstrated that AID and A1 are located within a cluster of pluripotency genes including *Nanog* and *Stella* on chromosome 6, and showed that all four genes were expressed in oocytes, ovaries, embryonic germ (EG) cells and embryonic stem (ES) cells, implying a role for deamination in early development. Further analysis revealed low levels of AID expression, but no detectable A1, in primordial germ cells (PGC), indicating that AID may play a role in removing methylation marks early in spermatogenesis. AID is also expressed later during spermatogenesis (Schreck *et al.*, 2006).

Scientists have long searched for an enzyme capable of actively removing the epigenetic 5-methylcytosine mark from DNA, so it was intriguing when AID was shown to be capable of performing this through deamination. AID deamination would result in a dG-dT mismatch, which would require further repair by a glycosylase enzyme such as thymidine DNA glycosylase (TDG) or methyl-binding domain protein 4 (MBD4). Rai *et al* presented exciting data suggesting that overexpression of AID, A2A and A2B (Zebrafish homologues of A2) in Zebrafish embryo's can promote demethylation of an injected methylated DNA fragment when MBD4 is co-expressed (Rai *et al.*, 2008). However, this data raises several questions as they propose that AID is acting on dsDNA yet AID has not been shown to deaminate dsDNA despite extensive investigation. Further MBD4 knockout mice do

not show demethylation defects indicating there is redundancy or a different pathway occurring in mammals. Nevertheless this does present an interesting concept which warrants further exploration.

Continuing the investigation of AID as a demethylating enzyme Bhutani *et al* employed a novel system to study candidate demethylases and analysed AID in this context (Bhutani *et al.*, 2009). They fused mouse embryonic stem (ES) cells and primary human fibroblasts to create heterokaryons, and then analysed the methylation and transcriptional state of human genes with human-specific primers. Using this method they detected extensive demethylation and transcriptional activation of human pluripotency genes (*Oct4* and *Nanog*) within one day of fusion and in the absence of replication. Knockdown of AID severely reduced the levels of demethylation observed indicating AID plays an active role in the demethylation of pluripotency genes.

Taken together these studies indicate a novel function of AID cytidine deaminase in the removal of the 5-methylcytosine mark from DNA. The discovery that AID is expressed early in development indicates it may function during the first wave of demethylation that follows fertilization. Although a potential new role for A1 in 5-methylcytosine deamination later in development should not be ignored (Morgan et al., 2004). Further work is required to determine the co-factors required for this process, and whether there is redundancy within the AID/APOBEC family as AID-null mice are viable (Muramatsu *et al.*, 2000) and do not display gross abnormalities or develop tumours indicative of de-regulation of methylation. Overexpression of Dnmt1 methyltransferase causes embryonic lethality in mice (Biniszkiewicz *et al.*, 2002) as does loss of the *de novo* methyltransferases *Dnmt3a* and *Dnmt3b* (Okano *et al.*, 1999) indicating regulated methylation is essential for viability. However it is unclear whether more than one mechanism of demethylation occurs and therefore what the phenotype associated with loss of a candidate demethylase would be. Consequently a more detailed analysis of methylation status in AID knockout mice is required. To accomplish this Popp and colleagues employed a genome wide bisulphite next generation sequencing approach to analyse DNA methylation levels

throughout the genome in wild type and AID deficient primordial germ cells (Popp *et al.*, 2010). This revealed that AID deficient mice retained higher levels of DNA methylation in primordial germ cells, specifically within introns and intergenic regions as opposed to promoters, but also on LTR-transposable element sequences. Interestingly, this was restricted to PGCs and the effect was more pronounced in female PGCs indicating an analogous system may be operational in the male germline.

APOBEC4

Given the diverse endogenous and important antiviral activities attributed to the AID/APOBEC family the discovery of a putative novel member of this family is intriguing. Bioinformatic searches of nucleotide genome sequences (NCBI and Ensembl databases) using the human AID protein sequence as a query revealed an uncharacterized vertebrate protein family termed APOBEC4 (A4) (Rogozin *et al.*, 2005), which was conserved from humans to *Xenopus tropicalis*. Importantly, reciprocal searches using A4 as the query returned other AID/APOBEC family members suggesting these proteins are homologs.

Alignments revealed conservation of the characteristic zinc-coordinating deaminase domain motif (H/C)xE….PCxxC (where x indicates any residue), although A4 contained a four amino acid insertion between the conserved cysteine residues, creating the motif PCxxxxxxC. However, longer inserts between the conserved cysteine residues have been observed in functional deaminases within the CDA superfamily, see Figure 1.2. These include members of the ADAR and ADAT enzymes (Dance *et al.*, 2001), although the substrate for both of these enzymes is structured RNA indicating the extra amino acids may enable the accommodation of larger substrates in the active site. Secondary structure prediction indicated A4 has the characteristic conformation of alternating α-helices and β-sheets which is shared by members of the CDA superfamily, and also contains an extra α-helix that is specific to the AID/APOBEC family (Huthoff and Malim, 2005). A4 has an

extended carboxy-terminal region that contains a coiled-coil domain not present in any other family members.

Phylogenetic analysis revealed that A4 is closely related to A1, with emergence of these genes from a common AID/A2 ancestor estimated to have occurred before the amphibian-reptile divergence (Rogozin et al., 2005). Importantly, sequence analysis indicated that A4 is evolving at a similar rate to A1, which is faster than AID but slower than A3 family members. The fast rate of evolution of A3 genes is attributed to their antiviral role in the innate immune system. The *in silico* expression profile of the A4 protein revealed that expression is restricted to the testis, indicating a potential function in spermatogenesis.

Mouse spermatogenesis

Mouse spermatogenesis is a continuous process which begins in the early embryo, by 7.5 days post coitum (dpc) there is a specified population of approximately 45 primordial germ cells (PGCs). PGCs are diploid pluripotent precursors of the primary spermatogonia, derived from proximal epiblast tissue (Lawson and Pedersen, 1992). Following proliferation and migration to the genital ridges male PGCs enter mitotic arrest at 13.5 dpc. In PGCs the inherited methylation pattern present at single copy genes and imprinted loci at 10.5 dpc is rapidly erased by 12.5 dpc when PGCs enter the gonad (Hajkova *et al.*, 2002). A *de novo* wave of monoallelic sex-specific methylation occurs throughout spermatogenesis which is performed by the *de novo* DNA methyltransferases Dnmt3a and Dnmt3b (Kato *et al.*, 2007). PGCs remain in mitotic arrest until shortly after birth when mitosis continues to produce type A spermatogonia stem cells, and the process of spermatogenesis begins.

The first wave of spermatogenesis occurs synchronously and is completed by 30 days post partum (dpp). This can be divided into three distinct processes: mitosis, meiosis, and spermiogenesis, which occur in the seminiferous tubule, reviewed in (de Rooij, 1998). Primary spermatogonia (undifferentiated diploid germ cells) are

located at the basal lamina, surrounded by supporting somatic sertoli cells. Germ cells undergo multiple rounds of asymmetric mitotic cell divisions to preserve a population of primary spermatogonial germ cells and produce differentiated primary spermatocytes that enter the first round of meiosis at approximately 9 dpp. Primary spermatocytes are characterized by highly condensed chromatin, and migrate from the basal lamina to the central lumen of the seminiferous tubule. After meiosis I secondary spermatocytes undergo one more meiotic division to produce haploid round spermatids, which progress towards the lumen as they mature. Finally, round spermatids undergo spermiogenesis (beginning around 20 dpp and complete by day 30), a metamorphosis into mature elongated spermatozoa. Spermiogenesis involves condensation of the genome into a nucleus approximately 5% the size of a somatic cell, shedding of excess cytoplasm and formation of the sperm tail. During spermatogenesis the whole genome undergoes extensive chromatin remodelling as histones are exchanged for protamines, coupled with establishing the pattern of paternal epigenetic modifications, which produces a highly condensed nucleus (reviewed in (Rousseaux *et al.*, 2005)). The mature sperm nucleus is believed to be transcriptionally silent due to chromatin compaction.

Methylation during early development

In mammals addition of a methyl group to a cytosine residue in the dinucleotide motif CpG is associated with heritable transcriptional silencing. Allele specific methylation can confer parental imprinting patterns and regulates the process of X-inactivation, whereas promoter regions of retrotransposons are heavily methylated on both strands to prevention of mobilization of retrotransposons. Approximately 70% of CpG dinucleotides are methylated in the vertebrate genome and these levels remain fairly constant in differentiated somatic tissues, however methylation levels show dynamic changes in the germline and in the preimplantation embryo (reviewed in (Morgan *et al.*, 2005)). *De novo* and maintenance DNA methyltransferases are responsible for transferring a methyl group from an s-adenosylmethionine (SAM) donor onto cytosine residues, this group of proteins is described in (Klose and Bird, 2006). However, to date there are no known enzymes capable of removing the

methylation mark. Several hypotheses have been suggested but they remain to be conclusively proven. Interestingly, a recent paper demonstrated that siRNA knockdown of components of the RNA polymerase II Elongator complex (Elp1, 2, and 3) impaired paternal genome demethylation after fertilization (Okada *et al.*, 2010). No catalytic demethylation activity was described but the focus of the study was on the Elp3 subunit which contains a radical SAM domain indicating it can bind to methyl groups. A homologous Archaea Elp3 protein is able to cleave SAM *in vitro* to generate methionine and 5'deoxyadenosine, raising the possibility that this could be occurring *in vivo* during demethylation (Paraskevopoulou *et al.*, 2006). Another potential mechanism for genome demethylation is cytidine deamination of 5-methylcytosine to thymine, with an intriguing paper highlighting the role of AID in this process in PGCs [discussed earlier see the section on AID as a demethylase and (Popp et al., 2010)].

At fertilization the paternal genome is haploid and the DNA is densely packaged around protamines. The female pronucleus however, is diploid, packaged with histones and arrested in metaphase II. Following fertilization, the male and female genomes remain condensed in distinct pronuclei, which only fuse at the first mitosis division. During this time the male pronucleus undergoes a rapid exchange of protamines for histones, followed by a wave of active demethylation before DNA replication has occurred. Demethylation of the male pronucleus occurs within 4 hours of fertilization in the mouse embryo and can be visualised using a 5-methylcytosine antibody (Santos *et al.*, 2002). Not all sequences are demethylated though, paternally imprinted genes and IAP retrotransposons remain methylated whilst L1 retrotransposons are demethylated (Lane *et al.*, 2003).

Sheep and rabbits do not have genome-wide demethylation of the paternal genome following fertilization, indicating that development can occur in some species without this step. This can be used to dissect the contribution of paternal and maternal factors and to determine the mechanism of demethylation. Beaujean *et al*, used intracytoplasmic sperm sperm injection (ICSI) to deliver mouse sperm into a sheep oocyte and showed that the male pronucleus underwent a limited loss of

methylation, indicating there is a paternal factor present in the mouse sperm which is capable of demethylating DNA (Beaujean *et al.*, 2004). The reciprocal experiment introducing sheep sperm into mouse oocytes showed a gradual loss of methylation that appeared to coincide with the decondensation of the paternal genome which occurs during the protamine to histone exchange, implying there is also a maternal factor present in the mouse oocyte that contributes to DNA demethylation.

1.5 The Adenosine Deaminase that act on RNA (ADAR) Family

ADARs were originally described as an activity that could unwind dsRNA and modify it such that it migrates differently when electrophoresed on a native gel and could not reform a duplex. This activity was first observed in *Xenopus laevis* embryo extracts, where up to 50% of the adenosine residues in the dsRNA tested were deaminated to inosines (Bass and Weintraub, 1987, 1988).

The ADAR family are thought to have evolved from Tad1/ADAT1 by the acquisition of several dsRBDs. *Drosophila melanogaster* contains an ADAT1 homolog, with a deaminase domain showing 38% identity and 47% similarity to the deaminase domain from human ADAR2 (Keegan *et al.*, 2000). This shows the ADAT1 protein, which also edits an adenosine residue in tRNA, is more closely related to the ADAR family than to the yeast Tad1 family. Interestingly, recombinant dADAT1 is unable to edit adenosine residues in dsRNA as ADAR family members do, as it lacks dsRBDs.

There are four members of the ADAR family in mammals, depicted in Figure 1.5. ADAR1 contains two translation start sites giving rise to two distinct isoforms that differ in their subcellular localisation and have different substrates (Kawakubo and Samuel, 2000). The longer ADAR1 isoform is interferon inducible indicating a role in the immune response, whereas the shorter isoform is constitutively expressed. *Drosophila melanogaster* has only one *Adar* gene, which shows the highest homology to mammalian ADAR2 (Palladino *et al.*, 2000a). Expression of ADARs

1, 2, and 3 is highest in the CNS, where most of the known specific substrates have been identified. Another member of the family known as testis-expressed nuclear RNA-binding protein (TENR) is expressed in the testis; however it lacks crucial conserved residues involved in catalysis and is likely to be catalytically inactive (Connolly *et al.*, 2005)

.

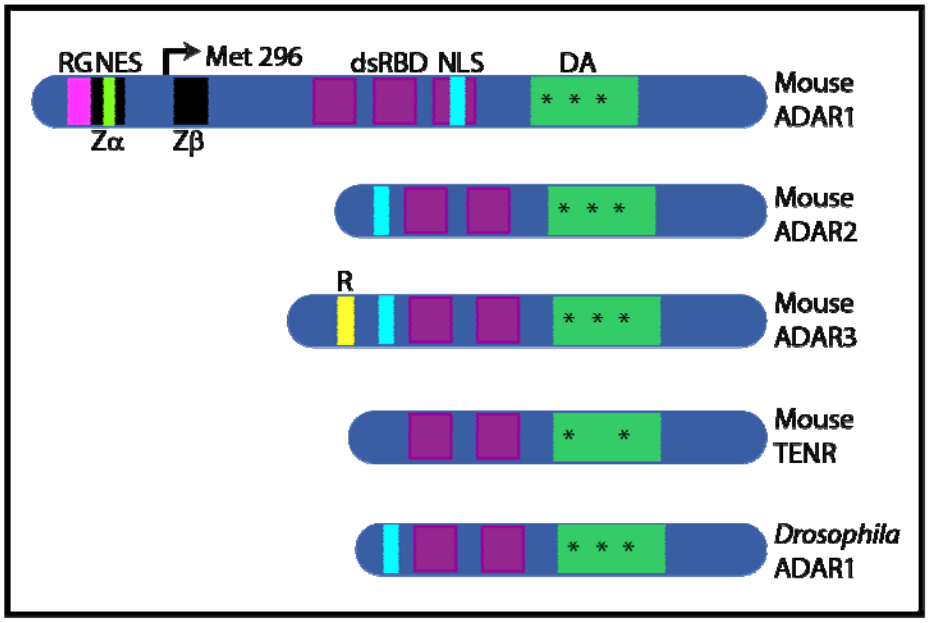


Figure 1.5: Cartoon depicting members of the ADAR family
A cartoon depicting the four ADAR family members in mouse: ADAR1, ADAR2, ADAR3 and testis-specific family member TENR which lacks a conserved residue within the deaminase domain. *Drosophila* ADAR has the greatest homology to mammalian ADAR2. The arrow indicates the alternative translation initiation site which produces ADAR1 p110 protein. DA indicates the deaminase domain, NLS indicates a nuclear localisation signal, NES indicates a nuclear export signal, dsRBD indicates the double-stranded RNA binding domains of which there are three in ADAR1 and two in other family members. RG indicates an arginine-glycine enriched domain, and R indicates an arginine enriched domain. The two Z-DNA binding domains in ADAR1 are labelled Zα and Zβ and asterisks indicate the three conserved motifs required for catalytic activity that are located within the deaminase domain.

ADAR1

The ADAR1 locus contains multiple promoters. Direct constitutive expression of transcripts from one promoter utilise an internal start site at methionine 296 to generate the p110 isoform (Kawakubo and Samuel, 2000). ADAR1 p110 protein was originally purified from calf thymus (O'Connell and Keller, 1994). Subsequently the cDNA was cloned (O'Connell *et al.*, 1995), and ADAR1 was

shown to edit the GluR-B transcript at the R/G site *in vitro* (Maas *et al.*, 1996). Mammalian ADAR1 p110 is a nuclear protein which is widely expressed, although higher levels are detected in the brain and CNS.

In the lymphatic system ADAR1 expression is regulated by an interferon-inducible promoter, which results in use of the first methionine to generate the longer p150 isoform (Patterson and Samuel, 1995). The p150 isoform is mainly cytoplasmic due to an NES signal within the first Z-DNA binding domain, leading to the hypothesis that this isoform may have additional cytoplasmic substrates from those of the constitutively expressed p110 nuclear isoform (Poulsen *et al.*, 2001). The Z-DNA binding domains allow interaction with nucleic acids in the Z-form such as DNA and structured RNA, and editing efficiency is lower when key residues in the Z-DNA binding domains are mutated (Herbert and Rich, 2001). ADAR1 p110 has been shown to localize to the nucleolus. Photobleaching experiments indicate this localisation is dynamic, and the protein leaves the nucleolus to interact with substrates (Desterro *et al.*, 2003).

A study into the genetics of Dyschromatosis Symmetrica Hereditaria (DSH) identified several mutations across the ADAR1 gene (Miyamura *et al.*, 2003). DSH is an autosomal dominant skin pigmentation disorder, which predominantly affects the hands and feet, although the disorder is relatively benign with no other symptoms described. The discovery of 30 different point mutations in the ADAR1 gene associated with this disorder indicates that ADAR1 has an as yet undefined role in melanocyte development or pigmentation. One reported case with the mutation G1007R also presented with chronic dystonic posturing and calcification in the brain. The mutated residue is highly conserved and lies within the deaminase domain. Functional assays show the G1007R mutant protein is not capable of RNA editing although it retains RNA-binding activity indicating it may behave as a dominant negative mutant (Heale *et al.*, 2009). This mutation has subsequently been reported in another individual with DSH, dystonia and calcinosis of the brain; however the mother who carried the mutation did not have neurological symptoms indicating this mutation may have reduced penetrance (Kondo *et al.*, 2008).

Altered RNA editing of the serotonin (5-HT2C) receptor transcript has been linked to schizophrenia and suicide (Niswender *et al.*, 2001; Sodhi *et al.*, 2001). However it is unclear from these studies on RNA editing levels in post-mortem brain whether the changes are causative of disease or a consequence of medication.

ADAR1-null mice

Unlike ADAR2 there is no transcript that requires to be edited by ADAR1. However, there are five known editing sites in exon 5 of the serotonin 5-hydroxytryptamine subtype 2C (5-HT_{2C}) receptor (Burns *et al.*, 1997) and ADAR1 has been shown to edit three of these sites *in vitro* (Liu *et al.*, 1999). It was therefore surprising when ADAR1-null mice displayed a severe embryonic phenotype of massive liver disintegration and died at embryonic day (E) 12.5 (Hartner *et al.*, 2004; Wang *et al.*, 2004). No role for ADARs during development has been described, and no edited transcripts in the liver have been found. Characterisation of the liver phenotype revealed a defect in proliferation of haematopoietic cells, with those present showing fragmented DNA indicative of apoptotic cell death.

Analysis of mouse embryonic fibroblast cells derived from ADAR1-null mice showed elevated apoptosis in response to serum starvation indicating a widespread predisposition to apoptotic cell death (Wang *et al.*, 2004). The onset of the ADAR1-null apoptotic phenotype correlated with an induction of the interferon-inducible p150 isoform following serum starvation implying a role for ADAR1 in mediating the induction of apoptosis in response to stress. Interestingly, analysis of the serotonin 5-HT_{2C} receptor transcript isolated from E12 ADAR1-null embryo's showed that exon 5, which contains all five editing sites, was aberrantly spliced out. This implies splicing and editing, or the lack of, are linked. Mice deficient in ADAR1 and ADAR2 show a similar phenotype to the single ADAR1 knockouts indicating ADAR2 does not have a similar role in early development (Hartner *et al.*, 2004).

Further characterisation of the liver phenotype in ADAR1-null mice revealed that ADAR1 is crucial for the maintenance of haematopoietic stem cells (Hartner *et al.*, 2009). During liver development ADAR1 acts to suppress interferon signalling and in the absence of ADAR1 elevated apoptosis occurs with aberrant activation of interferon-inducible genes. This is intriguing as ADAR1 is a dsRNA-binding protein and dsRNA is known to activate the interferon response (Geiss *et al.*, 2001). Therefore the hypothesis was proposed that an as yet unknown substrate of ADAR1 in the embryonic liver elicits an immune response in the absence of ADAR1 (Hartner *et al.*, 2009). The ADAR1-null phenotype could be rescued by expression of ADAR1 and rescue was largely, but not entirely, dependent on RNA-editing (XuFeng *et al.*, 2009).

ADAR2

The presence of inosine at the Q/R site in the GluR-B transcript, coupled with the observation that ADAR1 does not edit this site *in vitro*, led to the search for another deaminase and ADAR2 was identified (Melcher *et al.*, 1996b).

ADAR2 is a nuclear protein and like ADAR1 p110 it has been shown to localise to the nucleolus in a dynamic manner (Desterro *et al.*, 2003; Sansam *et al.*, 2003). Further investigation revealed that a region encompassing the first dsRBD is required for nucleolar localisation, and the ADAR2 protein re-localises to the nucleoplasm upon expression of a substrate. However, it is unclear whether the region harbours a nucleolar localisation signal or whether RNA-binding is required for the nucleolar localisation. Experiments with an ADAR2 mutant containing point mutations within both dsRBDs displayed exclusion from the nucleolus indicating RNA-binding was required for nucleolar localisation (Sansam *et al.*, 2003). These results suggest that ADAR2 activity is regulated through subcellular localisation.

The crystal structure of ADAR2 revealed an inositol hexakisphosphate molecule is buried in the core of the enzyme (Macbeth *et al.*, 2005). The residues which coordinate the inositol hexakisphosphate are conserved throughout the ADAR and

ADAT1 families, but absent from the ADAT2/3 family. The inositol hexakisphosphate molecule stabilises the ADAR2 protein structure and is essential for catalytic activity.

ADAR2 substrates

Glutamate receptors are grouped according to the synthetic agonists that they bind to, into α-amino-3-hydroxy-5-methylisoxasole-4-propionate (AMPA), N-methyl-D-aspartic acid (NMDA), and kainite receptors. Pre-mRNAs encoding AMPA and kainite receptor subunits are edited leading to variation at key residues within the receptor subunits. To date no mRNAs encoding NMDA receptor subunits have been found to be edited. NMDA receptors are highly permeable to calcium and mediate the slow component of the excitatory post-synaptic transmission. In contrast the AMPA receptors usually have low calcium permeability, although this is controlled by the presence of the GluR-B subunit which is edited, and mediate fast excitatory synaptic transmission.

Editing of the glutamate receptor B subunit transcript (GluR-B, also called GluR-2) changes a CAG codon to a CIG which is interpreted by the translation machinery as a CGG codon. This changes the genomically encoded glutamine at position 607 to an arginine residue, termed the Q/R site (Higuchi *et al.*, 1993). Editing of the Q/R site occurs to >99% in the mouse and human CNS. This editing event is performed by ADAR2, although ADAR1 can edit this site to approximately 10% as demonstrated in ADAR2-null mice (Higuchi *et al.*, 2000). Editing at this position regulates the permeability the AMPA receptor to calcium as inclusion of GluR-B subunits with the genomically encoded glutamine renders the channel permeable to calcium, whereas the arginine residue included as a result of RNA editing renders channels impermeable to calcium (Burnashev *et al.*, 1992). Thus, one site-specific RNA editing event has great functional importance for neuronal calcium homeostasis. However, AMPA receptors are tetrameric and can be formed by homo- or hetero-oligomeric assembly of any two or more subunits from a pool of four (GluR-A/B/C/D). Therefore receptors can lack a GluR-B subunit altogether making

them permeable to calcium, which occurs in aspiny neurons throughout the CNS. Calcium concentration is regulated in cells that do not express the GluR-B subunit through local calcium extrusion pumps and calcium-binding proteins. However, AMPA subunit composition is also dynamically remodelled in response to neuronal activity, the role of AMPA receptor subunits in synaptic plasticity is reviewed in (Liu and Zukin, 2007). Interestingly, despite sharing considerable sequence and structural homology the other three GluR subunit transcripts A, C, and D, do not undergo editing at the Q/R site and retain a glutamine at the corresponding amino acid.

Editing and alternative splicing of the GluR-B transcript has also been shown to affect receptor biogenesis in the endoplasmic reticulum (ER). Editing at the Q/R site results in retention of the GluR-B subunit with the ER, whilst the unedited GluR-B (Q) subunit is efficiently assembled into receptors and transported to the cell surface (Greger *et al.*, 2002). ER retention is thought to promote inclusion of the GluR-B (R) subunit into heterotetramers whereas GluR-B (Q) subunits more readily form homotetramers (Greger *et al.*, 2003). Interestingly it has recently been shown that RNA editing at the R/G site modulates AMPA receptor assembly in an analogous fashion by slowing the trafficking of AMPA receptors within the ER (Greger *et al.*, 2006). Therefore RNA processing events determine the kinetics and assembly of AMPA receptors at the synapse.

The calcium permeability of kainate receptors is also regulated by editing at one position which corresponds to the Q/R site in the GluR-B transcript. Editing of the transcripts encoding the kainate receptor subunits GluR-5 and GluR-6 is both spatially and developmentally regulated, and editing of each subunit transcript is independent of the other (Bernard *et al.*, 1999).

The transcripts encoding GluR subunits A-D are alternatively spliced which increases the variety of AMPA receptors. Two mutually exclusive exons termed FLIP and FLOP present in the ligand-binding domain of all four GluR subunits are thought to regulate the desensitisation kinetics of the channel (Koike *et al.*, 2000; Sommer *et al.*, 1990). Editing of another site in the GluR-B transcript which lies

close to the FLIP/FLOP exons results in a change from the genomically encoded arginine to a glycine residue (R/G site). The R/G site is also edited in GluR-C and GluR-D transcripts, however it is absent from GluR-A due to loss of a functional ECS region (Lomeli *et al.*, 1994). Editing at the R/G site is performed by both ADAR1 and ADAR2, and occurs at low levels during development increasing to approximately 75% in adult mouse brain. The resulting glycine substitution increases the rate of recovery of the channel from desensitisation (Lomeli *et al.*, 1994). Therefore the range of receptor subunits available is considerably increased through alternative splicing and editing.

ADAR2 transcriptional regulation

Rat ADAR2 is alternatively spliced which generates several different isoforms (Rueter *et al.*, 1999). One alternatively spliced exon located near the 5' end of the gene is 47bp long and inclusion of this exon alters the predicted reading frame of the protein. Use of the canonical translation start site in the presence of the 47bp exon would generate a truncated protein lacking the dsRBDs and the catalytic domain. However, use of an internal translation initiation site (Met25) generates a functional deaminase lacking the first 24 amino acids, although this protein is only detected at low levels in rat brain suggesting inefficient translation from the internal methionine. Further investigation revealed that the inclusion of this 47bp sequence was dependent upon editing of an adenosine residue at the -1 position of the 3' splice site, which would create an AI dinucleotide that mimics the canonical AG acceptor site. Editing at this site occurred to 47% in ADAR2 pre-mRNA isolated from rat brain. Therefore rat ADAR2 has evolved a complex autoregulatory feedback, such that alternative splicing of its own transcript is dependent on the presence of a functional ADAR2 protein (Rueter *et al.*, 1999). Mice lacking an ECS for the -1 editing position had elevated levels of ADAR2 protein and increased editing of ADAR2 substrates indicating the ADAR2 autoediting event functions *in vivo* to regulate ADAR2 activity (Feng *et al.*, 2006).

ADAR2 autoediting also occurs in human cell lines, where the level of auto-editing and inclusion of the alternatively spliced exon have been shown to correlate, such that higher expression of ADAR2 results in increased inclusion of the 47bp exon (Maas *et al.*, 2001). This is independent of ADAR1 expression. The stem-loop structure surrounding the edited splice site is created by interactions between intron 4 and the downstream exon and contains a loop of 1354 nucleotides, yet the paired regions are highly conserved. Interestingly, the adenosine residue at position -2 (neighbouring the -1 site) was shown to be edited by ADAR1 in cell culture experiments and editing occurred in 23% of ADAR2 transcripts isolated from rat brain indicating editing of this site occurs *in vivo* (Dawson *et al.*, 2004). Deamination of this residue would abrogate the cryptic splice site created by ADAR2 mediated editing of the -1 position, providing another potential mechanism to regulate ADAR2 activity via ADAR1.

Human ADAR2 cloned from a brain cDNA library showed two main isoforms, which differ through alternative splicing of an *Alu* cassette within the deaminase domain (Gerber *et al.*, 1997). Characterisation of the two ADAR2 isoforms revealed that they both showed the same substrate specificity, but inclusion of the *Alu*-containing exon led to a two-fold reduction in ADAR2 activity on a GluR-B R/G site substrate *in vitro*.

ADAR2 transgenic mice

In contrast to the embryonic lethal phenotype resulting from deletion of the ADAR1 gene, mice lacking ADAR2 survive to birth but died by post natal day 20 with an increasing predisposition to epileptic seizures (Higuchi *et al*., 2000). Interestingly, a similar phenotype was seen in mice engineered to express an uneditable GluR-B allele which lacks a functional ECS region (Brusa *et al.*, 1995), whereas mice that only express the edited version of GluR-B (R) are phenotypically normal (Kask *et al.*, 1998). The ADAR2-null mice show reduced editing at the GluR-B Q/R site: 10% of pre-mRNA transcripts were edited but this represents 40% of mature messages in these mice due to preferential splicing of the edited GluR-B transcript.

The authors attribute the low level of residual GluR-B editing to ADAR1 whose expression was unaffected by alterations in ADAR2. Further processing defects were observed for the GluR-B transcript. Inefficient editing led to accumulation of unspliced pre-mRNA in the nucleus that retained intron 11 where the ECS is located (Higuchi *et al.*, 2000). These processing problems resulted in a five-fold decrease in the amount of mature GluR-B mRNA. A similar problem with GluR-B processing was observed in mice expressing an uneditable GluR-B transcript lacking the ECS (Brusa *et al*., 1995). The RNA processing defects resulted in five-fold less GluR-B protein being produced, which led to an alteration in glutamate receptor physiology such that the CA1 pyramidal neurons displayed a 30-fold higher permeability to calcium compared to wild type. This defect in calcium homeostasis ultimately resulted in death of the affected neurons. The ADAR2-null phenotype could be completely rescued by expression of an edited GluR-B (R) transgene indicating the primary function of ADAR2 is to edit the GluR-B transcript (Higuchi *et al*., 2000). The rescue also confirms that there is no obvious role for the unedited GluR-B (Q) allele as the rescued mice were phenotypically wild type.

Interestingly, expression of a GluR-B transgene in mice with an asparagine residue at the Q/R site led to selective death of motoneurones that increased with age (Kuner *et al.*, 2005). The asparagine residue conferred a 2-fold increase in calcium permeability but did not significantly alter the AMPA channel conductance. The motoneuron phenotype was similar to that observed in a mouse model of familial amyotrophic lateral sclerosis (fALS) in which a mutation in the superoxide dismutase 1 (SOD1) gene results in selective motoneuron pathology (Gurney *et al.*, 1994). Indeed, crossing the GluR-B (N) mice with SOD1 mutants enhanced the motoneuron degeneration, implicating a role of GluR-B mediated calcium homeostasis in human motoneuron diseases such as ALS (Kuner *et al*., 2005). This is explored later in the section on ADAR2 as a "neuronal gatekeeper" and reviewed in (Buckingham *et al.*, 2008).

Transgenic mice overexpressing ADAR2 under the control of a CMV promoter developed adult-onset obesity, similar in phenotype to transgenic mice null for the

serotonin 5-HT_{2C} receptor (Singh *et al.*, 2007). Interestingly, the same adult-onset obesity phenotype was observed with transgenic mice overexpressing a catalytic site mutant ADAR2 E396A, indicating the phenotype was not due to aberrant adenosine deamination. Editing at several known sites including the serotonin 5-HT_{2C} receptor was shown to occur at similar levels to wild type littermates. However the transgenic mice also expressed endogenous ADAR2 and levels of transgenic ADAR2 protein in the brain were only 1.2-fold higher than endogenous levels, leading to the hypothesis that the phenotype was due to transgene expression in tissues that do not normally express ADAR2.

ADAR3

A third ADAR protein was initially identified in rat (Melcher *et al*., 1996a). The human homologue of this protein, ADAR3, is only expressed in specific regions of the brain including the amygdala and thymus (Chen *et al.*, 2000). *In vitro* studies of recombinant ADAR3 showed it was capable of binding to dsRNA like the other ADAR proteins, but also to ssRNA through an N-terminal Arginine and lysine rich region (R domain) that is not found in other members of the family. However, *in vitro* deaminase assays on known targets of other ADAR proteins (the GluR-B subunit transcript and the serotonin 5-HT_{2C} receptor subunit transcript) revealed that ADAR3 was not capable of deaminating these transcripts. A chimera generated from the ADAR3 dsRBDs and the ADAR2 deaminase (DA) domain was able to direct deamination of known substrates, but the reciprocal chimera containing the ADAR3 DA domain had no deaminase activity (Melcher *et al*., 1996a). *In vitro* analysis showed that ADAR3 could act as an effective competitor by binding to the same transcripts as ADARs 1 & 2 and preventing deamination (Chen *et al*., 2000).

To date no known function has been described for ADAR3, however the conservation of the deaminase domain as shown by amino acid alignment (Keegan *et al.*, 2004) indicates there may be selective pressure to keep it unchanged.

Testis Expressed Nuclear RNA-binding protein (TENR)

Testis-expressed nuclear RNA-binding protein (TENR), a novel member of the ADAR family, was found to be expressed in the round spermatids in the testis. TENR was identified in a screen to find RNA-binding proteins that interact with the 3'UTR of the *protamine 1* (*Prm1*) RNA (Schumacher *et al.*, 1995). TENR-null male mice are sterile (Connolly *et al*., 2005). This is due to low sperm counts and a high incidence of sperm morphological defects indicating TENR plays a role in sperm morphogenesis. However TENR lacks the conserved residues within the deaminase domain that are essential for catalytic activity indicating it is not a functional deaminase (Connolly *et al*., 2005). No substrates have been identified and the biological function of TENR remains unclear.

Drosophila ADAR

Drosophila melanogaster contains only one *Adar* gene and the dADAR protein shows the greatest homology to mammalian ADAR2, with the deaminase domain showing 67% identity to human ADAR2 (Palladino *et al.*, 2000a). To date several different *Drosophila Adar* mutant alleles have been described (Ma *et al.*, 2001; Palladino *et al.*, 2000b). Several deletion strains at the *Adar* locus were generated by P-element mobilisation, the largest of which encompasses the whole *Adar* gene and promoter region and is termed $Adar^{5G1}$. The ends of the $Adar^{5G1}$ deletion have been defined and the deletion also removes two neighbouring genes (*CG32806* and *CG14809*). These flies show a complete loss of both site-specific and non-specific RNA editing. $Adar^{5G1}$ flies also have severe behavioural and locomotive defects with no attempts at courtship and extremely uncoordinated movement with frequent falls. They display temperature-sensitive paralysis and tremors which progress in severity with age and flying is rarely observed and only occurs in response to repeated stimulation (Palladino *et al*., 2000b). $Adar^{1F1}$ mutants, in which only the promoter region of *Adar* is deleted, have reduced viability but under optimal environmental conditions they survive to 60 days, however in a mixed population the $Adar^{1F1}$ mutants fail to thrive indicating they are at a selective disadvantage. The ADAR-null flies developed a progressive age-dependent neurodegenerative

phenotype characterised by accumulation of vacuoles in the brain that was not observed in age-matched controls (Palladino *et al*., 2000b).

Similar behavioural defects to those observed in ADAR-null flies are observed in fly strains with mutations in ion channel genes which are known substrates of ADAR. For example flies with a temperature-sensitive mutation in the *cacophony* calcium channel display convulsions and with uncoordinated locomotion similar to *Adar* mutants. Defects in the courtship song were also observed for *cacophony*, *paralytic* and *slowpoke* temperature-sensitive mutants, and the transcripts from these genes are edited by ADAR (Peixoto and Hall, 1998). This indicates there is an overlap in the phenotype associated with *Adar* deletion and the defects arising from mutations in the genes which encode substrates of ADAR. However none of these mutants show a neurodegenerative phenotype indicating perhaps the neurological defects in *Adar*-null flies do not arise from defects in site-specific editing.

Two alleles of *Drosophila Adar* termed *hypnos-2*^{P} and *hypnos-2*^{L} were discovered in a genetic screen for loss-of-function mutations that show increase sensitivity to hypoxia (Ma *et al*., 2001). Ma *et al* showed that the *Adar* mutant *hypnos-2* had a delayed recovery following oxygen deprivation, and went on to characterise the *hypnos-2* mutant as having decreased locomotor activity and age-dependent neurodegeneration. The *hypnos-2*^{P} allele is a 434bp deletion of the genomic DNA encoding exons 5 and 6 of the *Drosophila Adar* gene. This is an in-frame deletion and removes the second dsRBD and part of the deaminase domain, yet the critical motif (HAE) containing the histidine residue that coordinates with the zinc atom and the glutamic acid residue that acts as a proton donor in the hydrolytic deamination reaction is not affected by this deletion. It is not clear whether protein is generated in these flies; however sequence analysis of known substrates of ADAR in the *hypnos-2*^{P} mutants revealed no evidence of site-specific editing. The *hypnos-2*^{P} flies have reduced locomotor function and reduced lifespan when compared to *Canton S* flies and develop progressive neuronal degeneration in the medulla and lobula complex. Furthermore the mutant phenotype could be rescued by expression of a wild type

Drosophila Adar transgene under control of the 32B promoter which expresses highly in the eye (Ma *et al*., 2001).

Drosophila ADAR expression

Drosophila Adar is expressed from 2 different promoters termed -4a and -4b which generate transcripts with different 5' exons. Whereas transcripts from the -4a promoter could be detected throughout development transcripts from the -4b promoter were only present in pupae and adult flies indicating this isoform is developmentally regulated (Palladino *et al*., 2000a). Transcripts from both promoters are subject to alternative splicing of exon -1, whereas only transcripts from the -4a promoter showed inclusion of the alternatively spliced exon 3a, giving rise to a total of 6 isoforms that are regulated both spatially and temporally (Marcucci *et al.*, 2009). Interestingly the alternatively spliced exon 3a occurs between the two dsRBDs. Altering the distance between dsRBDs could alter the substrate specificity of the deaminase as inclusion of exon 3a gives a protein that shows the same distance between dsRBDs as human ADAR2, and exclusion of exon 3a is comparable to human ADAR1 dsRBD spacing. Expression of *Drosophila Adar* is highest in the central nervous system throughout development, although expression is also detected in testis and ovary (Marcucci *et al*., 2009; Palladino *et al*., 2000a).

Sequence analysis revealed that the *dAdar* transcript undergoes auto-editing within exon 7 which produces an amino acid change of a serine to a glycine residue within the deaminase domain (Keegan *et al.*, 2005). Analysis of the *dAdar* transcript throughout development indicated that the auto-editing in exon 7 was highly regulated; editing was low in embryonic and pupal stages but increased after eclosion to reach approximately 40% in adult flies. However, further analysis revealed that transcripts containing the alternatively spliced exon 3a were never found to be edited, indicating that *dAdar* auto-editing is both transcript-dependent and temporally regulated (Palladino *et al*., 2000a). *In vitro* analysis revealed that the exon 7 editing site was edited to approximately 70% by a recombinant ADAR 3/4 S/G (lacking exon 3a, a mix of both edited and unedited isoforms) protein purified from *Pichia*

pastoris (Keegan *et al*., 2005). Analysis of the edited isoform containing a serine residue versus the unedited isoform containing a glycine residue, revealed that editing reduces catalytic activity by approximately 8-fold *in vitro*. Therefore dADAR editing of its own transcript provides an auto-regulatory feedback to lower catalytic activity. Indeed constitutive expression of an uneditable *dAdar 3/4 S* transgene under control of a ubiquitous actin promoter is lethal, whereas the glycine isoform is not (Keegan *et al*., 2005).

Characterisation of the role of ADAR within the developing nervous system revealed that activity increases throughout development with highest editing levels observed in the adult which correlates with increased auto-editing of *Adar* exon 7 (Keegan *et al*., 2005). Analysis of editing at 10 positions in the *cacophony* transcript demonstrated that editing frequency is low in embryos and highest in the adult with larval editing at an intermediate level. In *Drosophila melanogaster* many ADAR substrates have been identified indicating a more widespread role for dADAR in protein diversity. Sites in several substrates are edited to 100% however no transcripts have been identified that can rescue the ADAR-null phenotype as occurs with the GluR-B (Q) transcript in ADAR2-null mice (Higuchi *et al*., 2000). In *Drosophila* the compact genome restricts the number of proteins generated, therefore processes such as alternative splicing, alternative promoter usage and RNA editing are employed to increase the diversity of the protein repertoire.

RNA binding by ADARs

ADAR belongs to a diverse group of proteins that share one or more copies of a common dsRBD of approximately 70 amino acids (reviewed in (Fierro-Monti and Mathews, 2000)). Few residues are conserved throughout the family but all dsRBDs analysed share the same secondary structure of alpha helices and beta sheets, organised as α1-β1- β2- β3- α2, where the alpha helices make contact with the same face of the RNA. Other family members include *Staufen* a protein involved in mRNA transport; PKR, an interferon-inducible, RNA-dependent protein kinase; and

many proteins involved in the RNA interference pathway including Ribonuclease III, an RNA nuclease.

Importantly, these dsRBD-containing proteins do not exhibit sequence specificity, it is the secondary structure of the A-form RNA that they recognise. The substrate preference of RNA over DNA arises from the protein-RNA interactions that occur with the 2-hydroxyl group of the ribose sugar. Experiments to determine the binding and activity of ADAR2 were performed on a short RNA encompassing the GluR-B Q/R site and surrounding sequence to ensure secondary structure is formed (Stephens *et al.*, 2004). This demonstrated that ADAR2 dsRBDs exhibit selective RNA-binding on the GluR-B Q/R site and the regions contacted are distinct from the binding site of a dsRBD from PKR. The dsRBD occupies approximately 16bp of dsRNA but many can accommodate or prefer binding to bulges or loops within the dsRNA. However, individual dsRBDs display preferences for different duplex RNA structures, for example the ADAR2 dsRBDs exhibit different binding specificities when analysed separately which could enhance substrate specificity (Poulsen *et al*., 2006; Stefl *et al.*, 2006). dsRBD1 preferentially binds to perfect duplex dsRNA located in the stem-loop region, whereas dsRBD2 shows preference for dsRNA containing an A-C mismatch which is usually located near to the editing site. Both dsRBDs are required for efficient editing by ADAR2 (Stefl *et al*., 2006).

In mammals ADAR1 contains three dsRBDs whereas ADAR2 and ADAR3 contain two dsRBDs. *Drosophila* ADAR shows most homology to ADAR2 with two dsRBDs (Palladino *et al*., 2000a), which show the greatest homology to ADAR1 dsRBDs 1 and 3. ADAR1 dsRBDs 1 and 3 are required for catalytic activity, but dsRBD2 is dispensable, indicating the conserved dsRBDs 1 and 3 are involved in interactions with the RNA substrate (Lai *et al.*, 1995). Intriguingly, mutants analysed by Lai *et al*, that lack one of the conserved dsRBDs (ΔdsRBD1 or 3) do not show impaired binding to dsRNA but lose deaminase activity indicating binding and editing are separate events, and that binding of both dsRBDs is required for deamination (Lai *et al*., 1995). Analogous experiments with ADAR2 revealed that

deletion of dsRBD1 reduced editing efficiency whereas loss of dsRBD2 abolished editing altogether (Poulsen *et al*., 2006).

To investigate the specificity of the deaminase (DA) domain from ADAR1 and ADAR2 chimeric proteins with the dsRBDs from one ADAR and the DA domain from another were created (Wong *et al.*, 2001). This revealed that the editing specificity in the chimeric proteins was provided by the DA domain. In contrast, exchanging the dsRBDs between ADAR1 and PKR resulted in a loss of substrate recognition, indicating specificity is also provided by the dsRBDs although this chimeric protein retained the ability to deaminate some substrates (Liu *et al.*, 2000). ADAR3 is the only member of the ADAR family known to bind to ssRNA, and experiments have shown this interaction is mediated through the arginine and lysine-rich R domain in the amino terminus of the protein which is not found in other ADAR proteins (Chen *et al*., 2000). ADAR3 is also capable of binding dsRNA through its canonical dsRBDs but it is unable to edit either known substrates or dsRNA.

The formation of homodimers is required for adenosine deamination by ADAR enzymes (Cho *et al*., 2003; Gallo *et al*., 2003; Jaikaran *et al*., 2002; Poulsen *et al*., 2006). A ternary complex can be observed when increasing amounts of ADAR are added to substrate RNA, such that one monomer binds and then another indicating dimerisation is RNA-dependent (Jaikaran *et al*., 2002). Analysis of RNA editing *in vitro* using one wild type monomer and one catalytically inactive monomer showed that both monomers contribute to hyper-editing of dsRNA and site specific editing of substrates (Cho *et al*., 2003). Fluorescence energy resonance transfer (FRET) experiments indicate that ADAR1 and 2 form homodimers in an RNA-independent manner, and are capable of forming heterodimers *in vivo* (Chilibeck *et al*., 2006).

However, conflicting reports have been published on whether RNA-binding is required for dimerisation (Chilibeck et al., 2006; Gallo et al., 2003; Jaikaran et al., 2002; Poulsen et al., 2006; Valente and Nishikura, 2007). The conflict likely arises from the use of single or multiple point mutants which are predicted to abolish RNA-

binding but may have secondary effects on regulatory or structural residues which could account for differences in experimental results. Further, studies indicate that both dimerisation and RNA-binding are mediated through the dsRBDs, which complicates mutational analysis.

The minimal dimerisation domain for *Drosophila* ADAR comprises the N-terminal region including the first dsRBD, and mutations to abolish RNA binding within dsRBD1 also abolish dimer formation (Gallo *et al*., 2003). Yet a deletion which inhibits dimer formation does not affect RNA binding indicating monomers can bind dsRNA consistent with step-wise formation of dimers on dsRNA as observed for mammalian ADAR2 (Jaikaran *et al*., 2002). *In vitro* experiments mixing wild type dADAR with a mutant dADAR protein which lacks RNA-binding activity revealed editing efficiency was not affected indicating dADAR monomers do not interact without RNA-binding (Gallo *et al*., 2003). Similar experiments with a catalytic mutant dADAR demonstrated a 50% decrease in editing efficiency indicating both catalytic sites perform the deamination reaction. However for mammalian ADAR2, which shows the highest homology to *Drosophila* ADAR (Palladino *et al*., 2000a), the N-terminal region of the protein is dispensable for both dimerisation and editing (Macbeth *et al*., 2004; Poulsen *et al*., 2006). Addition of the N-terminal region of ADAR2 in *trans* reduced editing efficiency leading to the hypothesis that the N-terminal region of ADAR2 has an auto-inhibitory effect on catalytic activity. Binding of both dsRBDs to a dsRNA substrate is required to relieve the auto-inhibition for efficient editing (Macbeth *et al*., 2004).

To assess the role of RNA-binding in dimerisation Valente and Nishikura (2007) made mutations in three conserved lysine residues (KKxxK>EAxxA) within each of the dsRBDs of ADAR1 and ADAR2. The mutations introduced were based on data from alanine-scanning mutagenesis of the *Drosophila* RNA-binding protein *Staufen*, which demonstrated that mutation of exposed lysine residues within the conserved dsRBD eliminated RNA-binding without disrupting structure (Ramos *et al.*, 2000). Sequential purification of protein complexes containing both wild type and mutant ADAR proteins demonstrated that homodimerisation of ADAR proteins occurs

independently of RNA-binding (Valente and Nishikura, 2007). However dimeric ADAR proteins containing one mutant and one wild type monomer behaved in a dominant negative manner in both an RNA-binding assay and an *in vitro* editing assay indicating that RNA-binding of both monomers is required for deamination to occur. These results contradicts earlier studies utilising *Drosophila* ADAR protein which demonstrated that ADAR monomers do not interact without RNA-binding (Gallo *et al*., 2003). The differences likely arise from use of different mutations to abolish RNA-binding which may affect phosphorylation sites or dimerisation domains, alternatively the purification procedures may impact on the results as RNA may be present in the protein preparations. Interestingly Valente & Nishikura also observed that a dimer formed between a catalytic site mutant monomer and a wild type monomer exhibits normal RNA-binding and a 50% reduction in deamination *in vitro* (Valente and Nishikura, 2007), which confirms earlier results using dADAR (Gallo *et al*., 2003).

In yeast, ADAT2 and ADAT3 have been shown to function as a heterodimer where ADAT2 provides the catalytic function and ADAT3 provides substrate specificity (Gerber and Keller, 1999). This raised the possibility that ADARs may function as heterodimers. Heterodimerisation between ADAR proteins may reveal a role for ADAR3 where ADAR1 or 2 provides the catalytic activity and ADAR3 provides substrate specificity; however no heterodimer formation was found between members of the ADAR family in an *in vitro* analysis (Cho *et al*., 2003). This may be due to different subcellular localisation as the isoforms analysed were ADAR1 p150 which is cytoplasmic and ADAR2 which is nuclear. FRET analysis indicates heterodimers between ADAR1 and ADAR2 monomers form *in vivo* (Chilibeck *et al*., 2006) and heterodimers between ADAR1 and ADAR2 were co-immunoprecipitated from astrocytoma cell lines where it was demonstrated that increased levels of ADAR1 can inhibit editing of substrates by ADAR2 (Cenci *et al.*, 2008). It remains to be seen whether overexpression of ADAR3 can have the same effect. Dimers formed between the two isoforms of ADAR1 (p110 and p150) are readily detectable indicating that the Z-DNA binding domains absent in the p110 isoform are not required for dimerisation (Cho *et al*., 2003).

Secondary structure of RNA substrates

Initial investigation into editing of the GluR-B transcript at the Q/R site revealed an editing complementary sequence (ECS) located in the downstream intron was required for editing to occur (Higuchi *et al*., 1993). The ECS sequence forms an imperfect stem-loop structure which is bound by the dsRBDs of ADAR2. Point mutations which destabilised the stem-loop decreased the editing frequency, but this could be restored by making the complementary mutation in the opposite base, indicating it is the structure not the sequence that is important. Comparative SELEX using dsRBDs from *Xenopus laevis* ADAR1 and Xlrbpa revealed no sequence specificity but showed both dsRBDs had a similar structural preference for two stem-loop structures which incorporate bulges and mismatches, although perfect duplex RNA also showed strong binding affinity (Hallegger *et al.*, 2006).

Mammalian ADARs 1 and 2 show some overlapping substrates yet others are clearly specific for one or other protein, raising questions about substrate recognition. Whilst there is no consensus sequence surrounding edited adenosine residues ADAR enzymes exhibit local sequence preferences. For example: the 5' nearest neighbour preference for ADAR1 is U=A>C>G, similarly for ADAR2 it is U=A>C=G (Lehmann and Bass, 2000). However, editing does not occur at adenosine residues located near to the end of a dsRNA helix (Lehmann and Bass, 2000).

The carboxyl terminal region of ADAR enzymes contains motifs shared with DNA methyltransferase enzymes, which suggests that deamination of an adenosine residue may occur via a base-flipping mechanism similar to the addition of a methyl group by DNA methyltransferases (Hough and Bass, 1997). The addition of a methyl group by a DNA methyltransferase occurs outside the DNA helix, with the target residue rotated or flipped 180 degrees allowing modification of the internal face of the base and minimising structural disruption (Roberts and Cheng, 1998).

Substrates of mammalian ADARs

RNA editing plays a crucial role in the life cycle of the hepatitis delta virus (HDV) (Polson *et al.*, 1996). The viral genome contains an amber stop codon (UAG) which is altered by A-I editing to produce a UIG codon which is translated as tryptophan (UGG) by the translation machinery. The two proteins produced: short (p24) and long (p27), differ by 19 amino acids and have specific roles in the viral life cycle. p24 is required for replication of the viral RNA genome whereas p27 represses replication and is required for packaging of the virus. Editing at the HDV amber/W site was shown to occur *in vitro* with recombinant *Xenopus laevis* ADAR (Polson et al., 1996), and this *in vitro* editing event was used to analyse *Drosophila melanogaster* cell extract to show that an ADAR activity was present and conserved (Casey and Gerin, 1995). Editing at this site is inhibited by the HDV p24 protein via RNA-binding which provides an internal feedback loop that regulates the HDV life cycle through regulation of RNA editing (Polson *et al.*, 1998).

Inosine occurs at frequency of approximately one base in every 17,000 nucleotides in mouse brain poly $(A)^+$ RNA (Paul and Bass, 1998). This frequency of inosine nucleotides cannot be accounted for with known site-specific editing events and prompted the search for new ADAR substrates. Many attempts have been made to conclusively describe the list of ADAR substrates in different model organisms through detection of A-G transition changes when comparing cDNA and genomic DNA sequences. However this approach is difficult due to the high frequency of single nucleotide polymorphisms (SNPs) and sequencing errors.

Levanon *et al*, used a bioinformatic approach to search for expressed sequence tags (ESTs) containing A-G mismatches (Levanon *et al.*, 2004). As the substrate for ADAR is dsRNA they limited their search to double-stranded regions in ESTs. This approach identified 12,723 putative editing sites in 1,637 genes, 26 of which were validated by sequencing genomic DNA and cDNA from the same individual. Interestingly 92% of editing sites identified using this approach were located within *Alu* elements, and 1.3% were in LINE elements.

Using a similar approach Athanasiadis *et al,* searched databases for clusters of multiple A-G changes within short sequences, reasoning that these were unlikely to be SNPs or sequencing artefacts (Athanasiadis *et al.*, 2004). They also found that A-G changes clustered at the site of *Alu* repeat element insertions. This approach identified 1,445 mRNAs that were edited at 14,500 sites, which is equivalent to the majority of the inosine observed in brain mRNA. *Alu* insertion favours actively transcribed regions, often occurring in UTR regions and within intronic regions, therefore these regions are present in the pre-mRNA ADAR substrate. The edited *Alu* repeat elements were found to occur in tandem in inverted repeat orientation such that when transcribed a dsRNA stem-loop structure would form between the two inverted repeats and be edited by ADAR. Due to the extensive similarity between *Alu* elements the stem-loop structures formed are inherently stable, although bulges and mismatches were edited more frequently than near-perfect duplexes reflecting the *in vivo* substrate preference of ADARs. Some editing events were found to alter splice sites resulting in exonisation of a partial or whole *Alu* element, although the novel alternative splice sites were usually used at a low frequency and rarely constitutively (Athanasiadis et al., 2004). Surprisingly editing of other known repeat elements was considerably lower than that of *Alu* elements, which is likely to be due to the high level of conservation between *Alu* element sub-groups and the frequency of inverted insertions which create the required secondary structure.

The prevalence for editing within *Alu* elements was further confirmed using similar large scale bioinformatic approaches (Blow *et al.*, 2004; Kim *et al.*, 2004). These studies all conclude that the majority of pre-mRNA editing occurs within non-coding regions and may act to inhibit retrotransposition of repetitive elements. A further attempt to find novel editing sites in mammalian transcripts which generate re-coding events in the amino acid sequence used a screen based on known re-coding editing events (Clutterbuck *et al.*, 2005). This led to the identification of a novel ADAR substrate BC10, which contains three re-coding editing events.

Recently a deep sequencing project was performed to analyse potential editing sites in non-repetitive sequences using unique linked primers that form a “padlock” across

potential editing sites and following amplification and ligation these small regions were analysed by high throughput sequencing (Li *et al.*, 2009). This approach identified ten known substrates, along with 207 putative new substrates, although this list was considerable longer when the stringent selection criteria were relaxed as many transcripts showed low levels of editing (<5%). Of the sites identified 55 occurred in coding regions, and 38 led to amino acid changes, indicating re-coding editing events in humans are more common than originally thought. Editing also varied with brain region analysed, indicating there may be substrates of ADAR that are only edited in specific brain regions. Next generation sequencing technology has made the identification and verification of novel editing sites easier, however the functional consequences of these newly identified editing events remains to be determined.

Drosophila Adar substrates

RNA editing in *Drosophila* was initially described in the *rnp-4f* transcript isolated from fly head. However this editing was distinct from the site-specific editing observed in mice, as it appeared that numerous adenosine residues were edited throughout the transcript (263 A-G mutations in a 4kb mRNA) (Petschek *et al.*, 1996). This editing pattern is reminiscent of hyper-editing observed with perfect duplex dsRNA; however it was unclear how the dsRNA structure required for editing could form as the editing events occurred throughout coding and non-coding regions. Further investigation of the locus revealed that the neighbouring gene *sas-10*, which is transcribed from the opposite DNA strand, is transcribed from dual promoters. One promoter directs read-through of the termination signal to produce a long transcript which is perfectly complementary to the *rnp-4f* transcript. Interestingly the expression of these transcripts is inversely proportional such that *rnp-4f* is expressed early during development and declines into adulthood, and *sas-10* is induced at late stages of embryo development and peaks during the second instar larval stage. Sequence analysis revealed these RNAs are expressed in the same cells and both transcripts showed evidence of editing. However despite many attempts only one

cDNA was cloned which showed extensive non-specific hyper-editing indicating this is not a common occurrence (Peters *et al.*, 2003).

RNA editing has been described at four positions in the *paralytic* sodium channel subunit transcript (Hanrahan *et al.*, 2000). This provided the first comprehensive study of editing that resulted in re-coding in *Drosophila* and showed that editing of all positions was low in the embryo through to the third instar larvae stage and dramatically increased at pupation to high levels (maximum 70%) which were maintained into adult. Pairwise analysis of two sites revealed editing at each position was independent, and analysis of different isoforms generated by alternative splicing of one exon indicated that there was no preferential editing of specific isoforms. Comparison of *D.melanogaster* and *D.virilis* genomic sequence revealed highly conserved intronic regions likely to contain a putative ECS sequence, and one of these was confirmed by expression of a minimal cassette containing the edited site and the ECS region *in vivo* (Hanrahan et al., 2000). Interestingly, secondary structure prediction indicated that despite differences in the surrounding regions the edited site and the ECS from *D.melanogaster* and *D.virilis* formed almost identical stem-loop structures.

Maleless (*mle*) is a dsRNA helicase involved in sex-specific dosage compensation; however a *mle* mutant allele has been characterised that shows temperature-sensitive paralysis and was therefore called *no action potential* (nap^{ts}). The mle^{napts} mutant phenotype is due to a decreased level of correctly processed *para* transcript. *para* transcripts are aberrantly spliced in the region surrounding the editing events in the mle^{napts} mutant leading to the hypothesis that the dsRNA helicase is required to solve local secondary structures and allow correct splicing to occur (Reenan *et al.*, 2000). The mle^{napts} mutant probably acts as a dominant negative mutation by binding to the region but failing to unwind the RNA. After editing has occurred the duplex structure is usually resolved by an RNA helicase, however in the mle^{napts} flies this does not occur and a splicing catastrophe ensues. In support of this hypothesis the mle^{napts} mutant phenotype can be rescued by expression of an extra copy of *para*.

Further editing events were characterised in the *DmcalA* transcript which encodes a calcium channel subunit associated with the *cacophony* mutant (Keegan et al., 2005; Smith *et al.*, 1996), and the *GluClα* transcript encoding a chloride channel subunit (Semenov and Pak, 1999). Following publication of the *Drosophila* genome sequence further genes encoding nicotinic acetylcholine receptor subunits were identified and one of these (*nAcRα-30D*) was found to be edited at several sites in a developmentally regulated manner (Grauso *et al.*, 2002). The *nAcRα-30D* (also known as *Dα6*) transcript also undergoes alternative splicing which, together with editing of seven adenosine residues, can potentially generate >30,000 transcript variants. Since these discoveries several attempts have been made to characterise the complete range of ADAR substrates in *Drosophila*. As there is only one *Adar* gene in *Drosophila*, all identified editing events can be attributed to one enzyme.

The discovery that the sequence surrounding an editing site is protected from mutation allowed the use of comparative genomics to identify novel editing sites (Hoopengardner *et al.*, 2003). The secondary structure formed with the ECS is required for site-specific editing and mutations that disrupt the base pairing are not tolerated. Therefore comparing the genomic sequence of candidate genes from *Drosophila melanogaster* and *Drosophila pseudoobscura* to find highly conserved regions led to the identification of 16 novel ADAR substrates, these are listed in Table 1.2.

Category	Identifier	Gene Name	Function	Sites
Voltage-gated ion channels	CG34405	*NaCP60E*	Voltage-gated sodium channel	1
	CG15899	*Ca-α1T*	Voltage-gated calcium channel	1
	CG4894	*Ca-α1D*	Voltage-gated calcium channel	5
	CG12295	*straightjacket*	Voltage-gated calcium channel accessory subunit	3
	CG12348	*Shaker*	Voltage-gated potassium channel	6
	CG10952	*ether-a-go -go*	Voltage-gated potassium channel	6
	CG10693	*slowpoke*	Voltage-gated potassium channel	2
Synaptic release machinery	CG3139	*synaptotagmin 1*	Calcium sensing at synapse	4
	CG2999	*unc13*	SNARE binding	1
	CG12473	*stoned B*	Synaptic transmission	1
	CG32490	*complexin*	SNARE protein	3
	CG2520	*lap/like-AP180*	Synaptic transmission	1
Ligand-gated ion channels	CG4498	*nAcRα-34E*	Nicotinic acetylcholine receptor channel subunit	7
	CG11348	*nAcRβ-64B*	Nicotinic acetylcholine receptor channel subunit	4
	CG6798	*nAcRβ-96A*	Nicotinic acetylcholine receptor channel subunit	2
	CG10537	*Resistant to dieldrin*	GABA-receptor subunit	6

Table 1.2: ADAR substrates identified by Hoopengardner *et al,* 2003.
A total of 16 edited transcripts were identified using a comparative genomics approach to find conserved regions in candidate genes (Hoopengardner *et al*., 2003)

However, despite showing high levels of conservation across *Drosophiladae* species, ECS regions can be located up to 2 kilobases (kb) away from the site of editing. They can also form extremely complex secondary structures making identification if ECS's difficult. A well characterised example occurs in the *Drosophila melanogaster Synaptotagmin I* transcript which undergoes editing at four sites within the same exon. There are two separate ECS regions for these editing sites; both located over one kb downstream within an intron, which form a complex pseudoknot structure that spans several kb (Reenan, 2005).

Several approaches have been used to identify novel substrates for ADAR in different systems. Xia and colleagues (2005) developed an antibody against inosine which they used to enrich inosine-containing mRNA from *Drosophila* heads; this was then applied to a cDNA microarray (Xia *et al.*, 2005b). Using this approach a shortlist of 62 potential new substrates for ADAR were discovered, and further investigation of 12 transcripts revealed that 7 contained A-G mutations, these are listed in Table 1.3. Two of these transcripts have been previously identified as ADAR substrates; however known edited transcripts such as *para* and *cacophony* were not identified in this screen despite containing multiple editing sites and therefore multiple inosine residues per transcript. The authors attributed this to low

expression levels of these transcripts, and showed that the new substrates they identified were more abundant (Xia et al., 2005b). However the authors were also able to detect inosine-containing RNA in *Adar*-null flies used as a negative control. This could indicate that there is another mechanism for generation of inosine operating in *Adar*-null flies, possibly through spontaneous deamination, or it may indicate that the antibody used was not specifically recognising inosine. The fly strain used has a partial deletion of the *Adar* locus encompassing the second dsRBD and the beginning of the deaminase domain, and retains all of the conserved motifs within the deaminase domain indicating this strain may produce ADAR protein that retains some catalytically active *Adar* protein. This may explain how inosine was detected in these flies.

Identifier	Gene Name	Function	Sites
CG18314[$]	*DopEcR*	G-protein couples amine receptor	5
CG11348[$]	*nAcRβ-64B*	Nicotinic acetylcholine receptor channel subunit	2*
CG13167	-	Hydrogen-transporting two-sector ATPase	1*
CG9619	-	Protein phosphatase type 1 regulator	1
CG8428	*spinster*	Synaptic vesicle endocytosis, programmed cell death	1
CG12076	*YT521-B*	Unknown	2
CG14936	*Tetraspanin 33B*	Unknown	2

Table 1.3: ADAR substrates identified by Xia *et al*, 2005.
Two previously identified transcripts were confirmed and five new substrates were identified using an antibody to purify inosine-containing transcripts from fly head RNA (Xia et al., 2005b). Where [$] indicates transcripts previously identified * indicates transcripts that contained guanosine at the editing site in *Adar*-null flies.

More recently a list of 27 new ADAR substrates was published by Stapleton *et al*, (2006), listed in Table 1.4. They employed a different approach by analysing sequence data from the *Drosophila* Gene Collection database which includes cDNAs from over 10,000 genes (Stapleton *et al.*, 2006). Analysing sequence discrepancies in cDNAs from different tissues showed a clear predominance of A>G substitutions in head cDNA, indicative of RNA editing by ADAR. Eleven previously identified substrates were confirmed with this screening approach. This screen did not focus on candidate genes as previous screens have (Hoopengardner *et al*., 2003) which removes sampling bias, yet this screen also identified numerous transcripts involved in neurotransmission reflecting the high expression of *Adar* in the CNS.

Category	Identifier	Gene Name	Function	Sites
Vesicular traffic	CG34410	*Rab26*	GTPase activity	1
	CG11622	*Ral interacting protein*	Ral GTPase activator	10
	CG5627	*Rab3-GEF*	Rab guanyl nucleotide exchange factor	3
	CG14296	*endophillin A*	Promotes synaptic vesicle budding	2
	CG4260	*α-adaptin*	Component of endocytosis, subunit of AP-2	1
	CG8110	*sundaydriver*	Kinesin-dependent axonal transport	1
Ion homeostasis	CG4795	*Calphotin*	Calcium sequestration	1
	CG18660	*Nckx30C*	Potassium-dependent sodium/calcium antiporter	1
	CG1090	-	Potassium-dependent sodium/calcium antiporter	3
	CG5670	*Atpα*	Sodium/potassium exchanging ATPase	1
	CG32699	-	Calcium-binding, acyltranferse activity	1
	CG34123	-	kinase activity, transmembrane ion transport	5
Signal transduction	CG11711	*mob2*	Activator of Trc kinase	1
	CG8285	*bride of sevenless*	G-protein-coupled receptor	2
Ion channel	CG10706	*SK*	Small conductance calcium-activated potassium channel	2
	CG31116	-	Chloride channel	4
Cytoskeletal components	CG10076	*spire*	Actin-nucleation factor	1
	CG5166	*ataxin2*	Regulation of actin-filament formation	2
	CG42540	-	Structural constituent of cytoskeleton	6
Other	CG32809	-	ATP-binding	1
	CG9528	*real-time*	phosphatidylinositol transporter	1
	CG14616	*lethal (1) G0196*	acid phosphatase	3
	CG3556	-	Cobalamin (vitamin B12) binding	1
Unknown	CG1552	-	-	1
	CG31531	-	-	1
	CG9801	-	-	1
	CG12001	-	-	1

Table 1.4: ADAR substrates identified by Stapleton *et al*, 2006.
27 novel ADAR substrates were identified by screening of the *Drosophila* Gene Collection database and experimentally verified (Stapleton et al., 2006).

The total number of transcripts identified as ADAR substrates in *Drosophila* has now reached 55 and new questions have arisen about the biological function of these site-specific editing events. Recent advances in sequencing technology have enabled ideantification of hundreds of transcripts which undergo RNA editing in humans (Li *et al*, 2009), however it remains to be seen whether these new techniques will identify more substrates of *Drosophila* ADAR. To date most of the known ADAR substrates are involved in neurotransmission, whether at the level of recycling the neurotransmitter or of regulating the channel conductance. This indicates RNA editing plays a crucial role in modulating neural signalling, whether this is constitutive mechanism of generating protein diversity or regulated in response to different stimuli remains to be determined.

Consequences of Inosine-containing RNA

Deamination of adenosine bases in RNA results in transcripts that contain one or many inosine bases, which can have consequences for the cell. In coding sequences the inosine base is read as guanosine by the translation machinery which can lead to recoding of proteins with different properties as is the case with the GluR-B subunit, as described earlier. Deamination of adenosine residues that occur close to splicing junctions can affect the inclusion/exclusion of either introns or exons which can alter the RNA transcript. For example, the ADAR2 auto-editing site which changes a splice site and generates a non-functional protein (Rueter et al., 1999). In one example editing of the *PTPN6* transcript within an intron was shown to alter the conserved adenosine residue found at the branch point causing retention of the intron (Beghini *et al.*, 2000). The retained intron contains an in frame stop codon which is predicted to generate a truncated protein lacking catalytic activity. Editing can also remove a stop codon altering the protein produced as occurs with the HDV viral proteins p24 and p27 also described earlier (Polson et al., 1996). Additionally, editing can result in the exonisation of *Alu* elements via generation of splice site consensus sequences (Athanasiadis et al., 2004).

Editing of non-coding regions is widespread (Li et al., 2009), however little is known about the effects of these editing events. Editing within 3'UTR regions can alter miRNA binding sites (Borchert *et al.*, 2009) or polyadenylation signal sequences (Gu *et al.*, 2009) however this has yet to be described *in vivo*. Editing can also alter the processing of miRNAs or alter their target specificity. These processes are discussed later in the section on Editing and RNA interference.

The formation of one or many I-U base pairs in dsRNA can have structural implications as an I-U base pair is less stable than an A-U pairing, and has a considerably lower melting temperature (Serra *et al.*, 2004). I-U base pairs are also likely to alter the stacking of the dsRNA helix. In hyper-edited substrates these changes are likely to have a more substantial effect.

Hyper-editing of dsRNA transcripts where up to 50% of the adenosine residues in a single transcript are deaminated to inosines has been described for a voltage-dependent potassium channel (sqKv2) RNA from squid, where up to 17 adenosines are modified in a 360 base region (Patton *et al.*, 1997).

Highly edited RNAs were found to be cleaved by a cytoplasmic factor. Subsequent investigation revealed the Tudor staphylococcal nuclease (Tudor-SN) protein, a component of the RNA-induced silencing complex (RISC), specifically binds to inosine-containing RNA and promotes cleavage (Scadden, 2005). The cleavage site (5'-IIUI-3') was shown to be generated in dsRNA *in vitro* through editing by ADAR1, ADAR2 and *Drosophila* ADAR indicating that this is a highly preferred editing site (Scadden and O'Connell, 2005).

Inosine-containing RNA regulates gene expression in trans

Immobilised inosine-containing dsRNA was used to isolate proteins that specifically interact with inosine-containing dsRNA (Scadden, 2005). This led to the identification of numerous factors which are associated with stress granules (SG), including poly(A) binding protein (PABP), and components of the eukaryotic initiation complex (eIF-4A, 4G, and 4E) (Scadden, 2007). Cytoplasmic SG are formed when cells undergo environmental stress and proteins required for survival must be synthesised rapidly whilst synthesis of other non-essential proteins is halted, (reviewed in (Anderson and Kedersha, 2009)). As such SGs comprise a host of RNA processing proteins and cellular mRNAs that are maintained in a translationally silent state until the stress is relieved. The observation that SG-associated proteins also interact with inosine-containing dsRNA led to the hypothesis that inosine-containing dsRNA could affect mRNA expression levels through the induction of SGs. This hypothesis proved to be true for several different dsRNAs containing I-U base pairs and to a lesser extent dsRNAs with G-U base pairs, including an endogenously edited miRNA-142. Inosine-containing dsRNA transcribed within the cell resulted in a decrease in mRNA levels of both endogenous mRNAs and reporter mRNAs. This was accompanied by a decrease in translation which was due to inhibition of

translation initiation. A model was proposed whereby editing of dsRNAs by ADAR produces inosine-containing dsRNAs which are sequestered into SGs with cellular mRNAs and translation initiation factors leading to a decrease in translation within the cell (Scadden, 2007). The decrease in available translation factors results in increased traffic of cellular mRNAs to SGs, which are then either degraded in processing (P) bodies or held in a translationally silent state. The translation block is relieved when inosine-containing dsRNAs are degraded by Tudor-SN (Scadden, 2005) a component of the RISC complex, providing a mechanism for temporary silencing of endogenous mRNAs in response to dsRNA. However further work is required to test this model.

Nuclear retention of inosine-containing RNAs

Mouse CTN-RNA is a nuclear poly(A)+ transcript which is expressed highly in mouse liver, brain, and lung. It is expressed from the same locus as the mCAT2 transcript through use of an alternative promoter, which results in an almost identical transcript apart from a different 3'UTR (Prasanth *et al.*, 2005). The unique CTN-RNA 3'UTR is approximately 4.5kb, therefore the total transcript is approximately 8kb, compared to the cytoplasmic mCAT2 mRNA which is approximately 4.2kb. Within this long 3'UTR are four repeat sequences of approximately 100bp, predicted to have originated from SINE elements. The repeats are orientated such that one is forward and three are in reverse, giving the possibility of creating a large dsRNA hairpin structure which provides a substrate for ADAR. Analysis of transcripts from mouse liver revealed that the CTN-RNA 3'UTR was indeed edited at several positions within the repeat regions. Further investigation revealed that the edited CTN-RNA is retained within the nucleus, until cells are stressed whereupon the 3'UTR is cleaved and the CTN-RNA is released into the cytoplasm and translated to produce mCAT2 protein (Prasanth et al., 2005). Therefore, in this situation the unique CTN-RNA 3'UTR has a specific secondary structure, which is the target of editing by ADAR, and the transcript is retained in the nucleus until it is needed in times of stress. However this may be coincidental as no experiments were performed in the absence of ADAR therefore it is not clear what role, if any, ADAR plays in

altering the subcellular localisation of CTN-RNA. Further studies into human and *C.elegans* transcripts containing a dsRNA structure have failed to replicate these findings, in fact here the edited transcripts could be amplified from polysomes following sucrose gradient sedimentation indicating nuclear export and translation of edited transcripts occurs (Hundley *et al.*, 2008).

Editing and RNA interference (RNAi)

RNA interference (RNAi) is the process by which a double-stranded short interfering RNA (siRNA) or micro RNA (miRNA) of approximately 21-23 nucleotides (nt) can target the degradation of an mRNA with complementarity to it. The process of RNA silencing via siRNA and miRNA and their biogenesis is reviewed in (Carthew and Sontheimer, 2009). Briefly, siRNAs are cleaved from longer dsRNAs by DICER in the cytoplasm and show complete complementarity to their target mRNAs. Whereas miRNAs are processed from transcripts, often encoded in intronic regions of mRNAs, which form stem-loop dsRNA structures. These stem-loop primary miRNAs (pri-miRNAs) are cleaved from surrounding sequence in the nucleus by the microprocessor complex to generate a pre-miRNA. The pre-miRNA of approximately 75nt is then exported form the nucleus and processed further by DICER to remove the loop region and produce a mature miRNA (Kim, 2005). In contrast to siRNAs, miRNAs can contain mismatches although a critical "seed" region of 4-7nt must be complementary to their target mRNA. One strand of the mature miRNA or siRNA associates with the RISC complex (Gregory *et al.*, 2005) to direct silencing of target mRNAs, which can be mediated through cleavage and degradation of the target or translational inhibition. siRNAs and miRNAs activity depends on the formation of dsRNA during biogenesis and for functional RNA silencing, and given that ADARs bind dsRNA it has long been hypothesised that they could modulate the RNAi pathway.

An initial survey to analyse the extent of miRNA editing revealed that 13% pri-miRNAs were edited and that levels of editing varied across tissues analysed (Blow *et al.*, 2006). However this corresponded to 6% of mature miRNAs which were

edited and target site prediction software indicated that editing of mature miRNA could change their target specificity. A more comprehensive sequence analysis of primary miRNA (pri-mRNA) transcripts revealed numerous editing sites and led to the prediction that up to 16% pri-miRNAs could be edited in human brain (Kawahara *et al.*, 2008). This indicates that editing of miRNA occurs more frequently than initially thought.

The reported consequences of miRNA editing are varied, however only one example of re-direction of target specificity has been reported. In the case of the human miRNA-376 cluster editing occurred to nearly 100% at several sites (Kawahara *et al.*, 2007b). One editing site was located within the seed region of the mature miRNA and editing at this site was shown to alter the target specificity of the miRNAs produced. Furthermore the functional consequence of redirection of one mi-RNA was confirmed as altered levels of target protein expression and function were detected in ADAR2-null mice.

However, most reported cases of miRNA editing appear to inhibit biogenesis of the miRNA by DROSHA or DICER (Heale et al., 2009; Kawahara *et al.*, 2007a; Yang *et al.*, 2006). Although one report indicates this inhibition is mediated through binding of ADARs to miRNAs (Heale et al., 2009), and another through the presence of inosine in the pri-miRNA which inhibited cleavage by DICER (Kawahara et al., 2007a). Yang and colleagues found that processing of pri-miR-142 by DROSHA was inhibited by editing of adenosine residues located close to the DROSHA cleavage site (Yang et al., 2006). Consequently, downstream DICER processing was also inhibited; however edited pri-miR-142 did not accumulate to high levels due to specific degradation of inosine-containing pri-miR-142 by Tudor-SN. Therefore editing of pri-miR-142 serves to reduce the level of mature miR-142 by targeting the edited miRNA for degradation. Higher levels of endogenous miR-142 were found in ADAR-null mice confirming that this mechanism is used *in vivo* to regulate miR-142 levels (Yang et al., 2006). Therefore it is likely that ADARs can have numerous effects through either binding to or editing miRNAs during processing. However,

only a subset of edited pri-miRNAs have been identified indicating ADARs target pri-miRNAs with specific dsRNA structure (Yang et al., 2006).

Redirection of miRNAs can also occur via editing within the 3'UTR of the target mRNA. Editing frequently occurs within non-coding regions (Li et al., 2009). Borchert et al recently published a bioinformatic screen analysing the impact of RNA editing within UTRs on miRNA:mRNA interactions (Borchert *et al.*, 2009). They found that a quarter of the editing sites analysed (over 3000 sites from 12,723 analysed) generated new miRNA target sites and a considerable number of target sites were also destroyed indicating editing may have a wide impact on miRNA binding sites. Cell culture experiments indicate redirection can occur through this mechanism (Borchert et al., 2009; Heale et al., 2009). However the difficulty in functionally verifying alterations in a miRNA target sequence has hampered investigation of this phenomenon *in vivo*.

Caenorrhabditis elegans contains two *adar* genes and strains lacking both *adr-1* and *adr-2* were used to investigate antagonism of RNAi (Knight and Bass, 2002). This revealed that *adr-1/adr-2* mutant worms showed strong somatic transgene induced RNAi which was absent from wild type worms. This is due to editing of the dsRNA in wild type worms which prevents the dsRNA from cleavage by Dicer and entry into the RNAi pathway. However *C.elegans adar* genes are not well conserved and editing events which lead to re-coding have yet to be identified in *C.elegans* indicating *adar* may have a different role in nematodes. Therefore *C.elegans* may provide an ideal system to investigate the editing-independent function of ADARs.

ADAR and disease

Several groups have reported a link between decreased levels of ADAR activity and tumour progression (Cenci et al., 2008; Maas et al., 2001; Paz *et al.*, 2007). Decreased editing of the GluR-B and serotonin receptor transcripts was first described in malignant gliomas and a correlation was shown between tumour stage and editing level indicating a progressive loss of editing (Maas et al., 2001).

However levels of ADAR remained unchanged indicating the change in activity was post-translational. Investigation on non-specific editing of *Alu* repeat sequences indicated significant global hypoediting in tumours from brain, prostate, lung, kidney and testis tumours (Paz *et al.*, 2007). Further analysis showed that RNA levels of ADAR-1, -2, and -3 were also reduced, and overexpression of ADAR-1 or -2 in a glioblastoma cell line resulted in decreased proliferation, leading the authors to conclude that the loss of editing enzymes was involved in the pathogenesis of cancer.

ADAR activity was also shown to be reduced in pediatric astrocytomas (Cenci et al., 2008). In this study the level editing was also shown to correlate with the grade of malignancy such that high grade tumours showed the lowest levels of editing. In astrocytoma cell lines the low levels of editing were restored by expression of ADAR2, and this inhibited proliferation of the tumour-derived cell lines. However, no difference in ADAR2 expression was detected between tumour and control samples. Investigation of the expression levels of other ADARs revealed both ADAR1 and ADAR3 were significantly overexpresses in the astrocytomas. *In vitro* experiments confirmed that increased expression of ADAR1 led to decreased editing of ADAR2 substrates. The ADAR1 and ADAR2 proteins were shown to co-immunoprecipitate indicating heterodimers were formed and this decreased editing of ADAR2 substrates (Cenci et al., 2008).

ADAR2 protects against neuronal degeneration following ischemic insult

ADAR2 has recently been implicated in recovery from forebrain ischemia. Following transient disruption of the blood flow to the brain, which can be caused by several factors including cardiac arrest, there is a selective degeneration of specific neuronal subsets. This has been observed for a long time, however it was unclear what determined the specific vulnerability of some neuronal subsets and not others. In the hippocampus one of the affected neuronal populations is the CA1 pyramidal neurons, which were shown to have an 18-fold increased permeability to calcium following ischemic insult (Liu *et al.*, 2004). Further investigation revealed the

increased permeability to calcium resulted from a decrease in edited GluR-B (R) which correlated with a decreased level of ADAR2 mRNA. This led to the "GluR-B hypothesis" that ADAR2-mediated editing of GluR-B Q/R was disrupted in CA1 pyramidal neurons following ischemic insult, and this led to an influx of calcium which led to neuronal degeneration. In support of "GluR-B hypothesis" Peng *et al*, (2006) showed siRNA knockdown of *Adar2* in the dentate gyrus increased the sensitivity of neurons to ischemic insult (Peng *et al.*, 2006). The loss of ADAR2 led to decreased editing at the GluR-B Q/R site, which increased AMPA receptor calcium permeability and led to neuronal death. Interestingly, the neuronal death associated with knockdown of *Adar2* could be rescued by expression of the edited GluR-B (R) transcript, in a similar manner to the rescue of the ADAR2-null mouse (Higuchi et al., 2000). Further ectopic expression of ADAR2 was shown to have a protective effect, and reinstate editing at the GluR-B Q/R site to >95%. Finally the authors showed that expression of ADAR2 is regulated by cyclic AMP response element binding protein (CREB) (Peng et al., 2006) which has been shown to be decreased in CA1 pyramidal neurons following ischemic insult (Walton *et al.*, 1996). Taken together these results implicate ADAR2 as an important effector of cell death, such that when stress pathways are activated ADAR2 levels fall, calcium permeability increases and neuronal degeneration occurs. However similar studies have failed to replicate the observation that editing levels are lower following ischemic insult, which could reflect analysis of different regions of the brain or large tissue samples which contain multiple cell types.

ADAR2 as a "neuronal gatekeeper"

A recent study showed that editing of the GluR-B subunit transcript at the Q/R site was downregulated in single motor neurons isolated from patients with sporadic amyotrophic lateral sclerosis (sALS) (Kawahara *et al.*, 2004). ALS is a progressive, degenerative disease characterised by a selective loss of motor neurons which leads to paralysis and death. There is currently no effective treatment. This led to the hypothesis that GluR-B acts as a "neuronal gatekeeper" in the same manner as the

model of ischemic stroke, such that decreased editing of the GluR-B Q/R site renders the neuron permeable to calcium which leads to cell death (Buckingham et al., 2008).

Editing of the GluR-B transcript was investigated as motor neurones are vulnerable to AMPA-receptor mediated neurotoxicity similar to CA1 pyramidal neurones as described above (reviewed in (Kwak and Weiss, 2006)). Editing at the Q/R position of the GluR-B transcript was analysed in single motor neurones isolated by laser-capture microdissection from the spinal chord of patients with ALS and controls. Editing efficiency was 100% in control motor neurones; however motor neurones form ALS patients showed a range of 0-100% editing, with an average of approximately 60% (Kawahara et al., 2004). The decrease in editing was specific to the motor neurones as editing in Purkinje cells from these patients was 100%, and it was also specific to sALS patients as patients with another neurodegenerative disease retained 100% editing. The decrease in GluR-B Q/R editing correlated with a decrease in *Adar2* RNA levels (Kwak and Kawahara, 2005). The observation that GluR-B-deficient mice do not develop seizures despite showing increased calcium permeability indicates that having unedited GluR-B present is more toxic to neurons than absence of the GluR-B subunit (Jia *et al.*, 1996). However, loss of GluR-B Q/R site editing is not observed in motor neurons from patients with familial ALS (fALS) or mouse models of fALS that contain mutations in the SOD1 gene, indicating despite similar phenotypes there are multiple mechanisms of pathogenesis of ALS.

Drosophila models of human neurodegenerative disorders

Drosophila melanogaster is an ideal model organism for studying processes that are involved in human neurodegenerative disorders. The short lifespan, ease of genetic manipulation, and wealth of anatomical information about the CNS along with the high conservation of cell biological pathways makes *Drosophila* a valuable tool for investigating neurological disease processes. *Drosophila* display complex motor coordination and courtship rituals, indicating that they are capable of complex

processing and memory functions. Coordinated locomotion provides an easily quantified measure of complex brain function.

Drosophila mutants for modelling human neurodegenerative diseases fall into two general classes (reviewed in (Muqit and Feany, 2002)). Firstly, transgenic *Drosophila* mutants that are null for a homolog of a disease-causing gene or express a human gene associated with neurodegenerative disease, or a mutant thereof. These models require prior knowledge of disease-associated genes but can be used effectively to recreate aberrant expression patterns associated with human disease. Examples of these models include flies overexpressing polyglutamine expansions, which can be used to model human disorders including Huntington's disease (HD). Fly models of Huntington's have been used to demonstrate that expression of an expanded polyglutamine peptide alone cause neurodegeneration (Marsh *et al.*, 2000). This observation highlighted the toxicity of the gain-of-function polyglutamine expansion whilst revealing the contribution of the protein containing the polyglutamine expansion to the disease process. Although the pathology is similar across polyglutamine expansion disorders, specific neuronal subsets are affected in different disorders, which is thought to be due to the protein containing the expansion.

The second type of *Drosophila* model has mutations in endogenous genes which produce a neurodegenerative phenotype. This approach requires no prior knowledge of the gene involved and can be used to identify genes involved in neurodegeneration that may otherwise remain unknown. Examples of this include the *Drosophila* mutants *swiss cheese* (Kretzschmar *et al.*, 1997) and *bubblegum* (Min and Benzer, 1999). *swiss cheese* (*sws*) mutants develop age-dependent neurodegeneration with extensive apoptosis and an abnormal glial phenotype (Kretzschmar et al., 1997). Further work has shown that the *sws* homolog in mouse also produces a neurodegenerative phenotype when mutated, and the SWS protein is thought to function in the metabolism of organophosphates, indicating it could be functionally important in organophosphorus-induced delayed neuropathy (Muhlig-Versen *et al.*, 2005). The *bubblegum* mutant was isolated in a screen to identify P-element

mutations that produce a neurodegenerative phenotype (Min and Benzer, 1999). *bubblegum* mutants were shown to have elevated levels of very long chain fatty acids (VLCA), and the causative mutation was identified in a VLCA acyl coenzyme A synthetase-like gene. This phenotype was similar to a human disease adrenoleukodystrophy (ALD), in which causative mutations have been mapped to the ABCD1 peroxisomal transporter. VLCA levels are altered in ALD and *bubblegum* could now become a model for screening therapeutics for ALD.

Suppression of neurodegeneration by expression of Adar in the cholinergic neurons

A previous student in the lab Leeanne McGurk characterised the neurodegenerative phenotype in $Adar^{5G1}$ flies as a progressive vacuolisation of the neuropil in the mushroom bodies with associated retinal degeneration (McGurk. L., 2008). This phenotype can be suppressed by expression of an *Adar* transgene in the cholinergic neurons utilising the binary expression system developed by Brand & Perrimon (Brand and Perrimon, 1993). $Adar^{5G1}$ flies with GAL4 expression in certain cell subsets, here cholinergic neurons using the *Cha-GAL4* driver, were crossed with flies that contain a randomly integrated transgene which comprises the GAL4 upstream activating sequence (UAS) directing expression of a gene of interest. Using this system to express *Adar* in the cholinergic neurons restored locomotion defects observed in $Adar^{5G1}$ flies indicating the phenotype is due to loss of ADAR function. Expression of a catalytically inactive *Adar* also showed suppression of the neurodegenerative phenotype but interestingly this did not restore the observed locomotion defects.

Further investigation demonstrated that human ADAR1 p110, ADAR2 and ADAR3 could suppress the neurodegenerative phenotype however ADAR1 p150 could not, indicating expression of ADAR in the nucleus is critical for suppression of the $Adar^{5G1}$ phenotype (Leeanne McGurk, Thesis). Although suppressed by expression of these *Adar* transgenes evidence of the neurodegenerative phenotype was still present, seen as holes in the brain, but fewer in number and smaller than those

observed in the *Adar*^{5G1} flies. This indicates the progressive nature of the neurodegenerative phenotype, and suggests that the suppression may act to delay the onset of degeneration indicating it is not a complete "rescue".

Comparison of GAL4 driver lines revealed that constitutive expression of *Adar* using an *Actin* promoter was lethal, indicating high levels of *Adar* are not tolerated. Other driver lines were tested for their ability to rescue the *Adar*^{5G1} locomotion defects. The *D42-GAL4* strong neuronal driver (Parkes *et al.*, 1998) was able to suppress the locomotion defects, as was a muscle tissue driver *Mef2-GAL4* although this also directed expression to the mushroom bodies (Ranganayakulu *et al.*, 1996). Neither of these lines was able to restore locomotion to wild type levels and *Cha-GAL4* gave the best rescue results (Keegan et al., 2005).

1.6 Alternative splicing

Splicing of pre-mRNA is the process of removal of intronic sequences and splicing together exonic sequences to form a mature mRNA and was initially described in the late 1970's (Berget *et al.*, 1977; Chow *et al.*, 1977). Splicing occurs in the nucleus when short, degenerate sequences which demarcate the intron-exon boundaries are recognised by splicing factors; see Figure 1.6 for a summary. Splicing is performed by a large multiprotein complex termed the spliceosome which assembles at the splice site and catalyzes the two-step splicing reaction (for a review see (Black, 2003)). Briefly, small nuclear ribonucleoproteins (snRNP) components of the spliceosome complex recognise the consensus sequences at the 5' and 3' splice sites, the branch point is bound by the SF1 branch point binding protein and the U2 auxilliary factor 65kDa subunit (U2AF65) bind to the polypyrimidine tract. In the first reaction the 5' splice site of an exon performs a nucleophilic attack on the conserved adenosine residue located within the branch point (BP) consensus sequence (YNYURAY). This results in cleavage of the exon from the intron at the 5' splice site and ligation of the 5' end of the intron with the branch point adenosine creating a lariat structure. In the second reaction additional snRNPs are recruited and the free hydroxyl group at the 5' exonic splice site performs a nucleophilic attack on the 3' splice site and the resulting transesterification reaction fuses the 5' and 3'

splice sites of the exons together and releases the intronic lariat sequence (Moore and Sharp, 1993).

The consensus sequence which demarcates the 5' splice site is A/CAG|GURAGU and that which denotes the 3' splice site is YAG|R, where | indicates the exon-intron boundary, R indicates any purine, Y indicates any pyrimidine, and the underlined nucleotides are invariate. Both these sequences contain adenosine and guanosine residues, of particular note is the invariate guanosine residues at both 5' and 3' splice sites. As inosine is read as guanosine editing of an adenosine residue which follows another adenosine would create an AI dinucleotide which mimics a 3' splice site. Similarly editing of the adenosine of an AG dinucleotide present at a 3' splice site would generate an IG dinucleotide which abolishes the splice site. Therefore there is a great potential impact of editing on altering splice site consensus sequences.

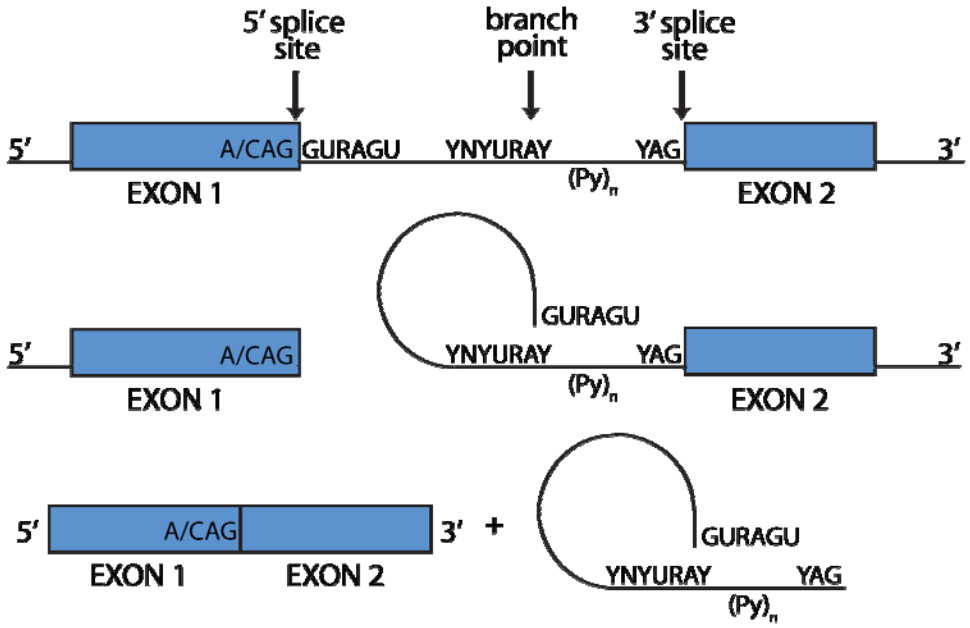


Figure 1.6: Pre-mRNA Splicing

Pre-mRNA splicing occurs co-transcriptionally and results in the removal of intronic sequence and the splicing together of exonic sequences. The intron-exon boundaries are demarcated by conserved sequence elements. The 5' splice site consensus sequence is A/CAG| GURAGU, whilst the 3' splice site is defined by the sequence YAG, where Y indicates any purine, and | indicates the exon-intron junction. The splicing reaction is performed by a multi-protein complex termed the spliceosome. Briefly, an intermediate "lariat" is formed when the 5' splice site performs a nucleophilic attack on the adenosine residue within the consensus *branchpoint sequence (YNYURAY), where R indicates any purine residue.* The free hydroxyl group at the 5' splice site then undergoes a second transesterification reaction with the 3' splice site resulting in fusion of the 5' and 3' splice sites and release of the intronic lariat (Moore & Sharp, 1993).

Many mRNAs have different splicing patterns and alternative splicing (AS) can generate multiple mature mRNAs from one pre-mRNA. Different patterns of AS, summarised in Figure 1.7, can be utilised at one locus to generate many different mRNAs with distinct functions (reviewed in (Black, 2003)). Regulation of AS can occur at the stage of spliceosome assembly, however there are numerous non-spliceosomal RNA-binding proteins such as the SR and hnRNP protein families which regulate AS through binding to regulatory sequences. Binding of accessory proteins can enhance or repress splicing of neighbouring exons and the abundance enhancer and repressor proteins likely plays a role in determining the splicing pattern. Control of AS can be sex-specific, developmentally regulated, tissue-specific or even controlled through signalling cascades such as the insulin response pathway (Hartmann *et al.*, 2009). A well documented AS cascade involves the somatic sex determination pathway in *Drosophila* where the female specific RNA-binding protein Sex Lethal (*sxl*) represses splicing patterns which lead to male development. This example of sex-specific splicing regulation is described in detail in (Black, 2003). However unlike the regulation of sex determination through AS which generates either a female or a male splicing pattern, many AS genes generate numerous transcripts from one locus and the balance between the isoforms generated can be crucial. AS can alter the reading frame of a transcript or cause the inclusion of a premature termination codon (PTC) which targets the transcript for nonsense mediated decay (NMD) (Lejeune and Maquat, 2005).

Levels of AS correlate with complexity of tissue and AS occurs at high levels in the human CNS creating splicing patterns which are distinct from other tissues (Yeo *et al.*, 2004). One example of the variation that AS can generate from one locus is the *Drosophila melanogaster Down syndrome cell adhesion molecule* (*Dscam*) gene. The *Dscam* pre-mRNA contains 115 exons which are organised into twenty constitutively spliced exons and four clusters of exons which are spliced in a mutually exclusive manner. The exon 4 cluster contains 12 alternative exons, the exon 6 cluster has 48 alternative exons, the exon 9 cluster has 33 alternative exons, and the exon 17 cluster contains 2 alternative exons. Assuming all combinations are possible the repertoire of isoforms generated from the *Dscam* locus is 38,016

different transcripts (Schmucker *et al.*, 2000). The alternative exons 4, 6, and 9 encode immunoglobulin domains whilst alternative exon 17 encodes a transmembrane domain.

Investigation into the mechanism of AS at the *Dscam* locus revealed that the mutually exclusive splicing of one exon from the exon 6 cluster is regulated through extensive secondary structure formation (Graveley, 2005). Interaction between highly conserved regions within the *Dscam* gene generates a stem-loop structure. A "docking site" located downstream of the constitutively spliced exon 5 base pairs with a "selector site" which is immediately upstream of each of the exon 6 alternatives, and ensures that only one of the mutually exclusive exons can associate with the 5' splice site. Heterogeneous nuclear ribonucleoprotein hrp36 ensures only one of the exon 6 alternatives is included in the mature *Dscam* mRNA through binding to all the exon 6 alternatives and repressing activity of SR proteins (Olson *et al.*, 2007).

DSCAM is involved in neuronal wiring and axon guidance, and multiple isoforms are required for axon self-avoidance through homophilic binding (reviewed in (Schmucker, 2007). The *Dscam* gene demonstrates the complex AS events which must be carefully regulated to ensure correct development of the *Drosophila* brain. Further investigation of the mechanism of *Dscam*-mediated self avoidance revealed that over 1000 isoforms are required for correct neuronal connections (Hattori *et al.*, 2009). Genetic manipulation to remove the multiple alternative exons from specific clusters showed that flies expressing 4,752 isoforms were indistinguishable from wild type and that the neurological defects associated with limited *Dscam* diversity get progressively more severe as the number of potential *Dscam* transcripts are limited.

Trans-splicing has also been demonstrated in *Drosophila melanogaster*. This occurs in a similar manner to pre-mRNA splicing but results in the fusion of two separate mRNAs to create a new transcript. Trans-splicing occurs at the *mod(mdg4)* locus, where the common constitutively cis-spliced N-terminal four exons are fused to one

of 26 different downstream regions, reviewed in (Dorn and Krauss, 2003). An unusual feature of this locus is that some of the downstream regions are transcribed from the same DNA strand as the common exons whilst others are transcribed from the antiparallel DNA strand, although all 26 isoforms are generated through trans-splicing of two separate transcripts.

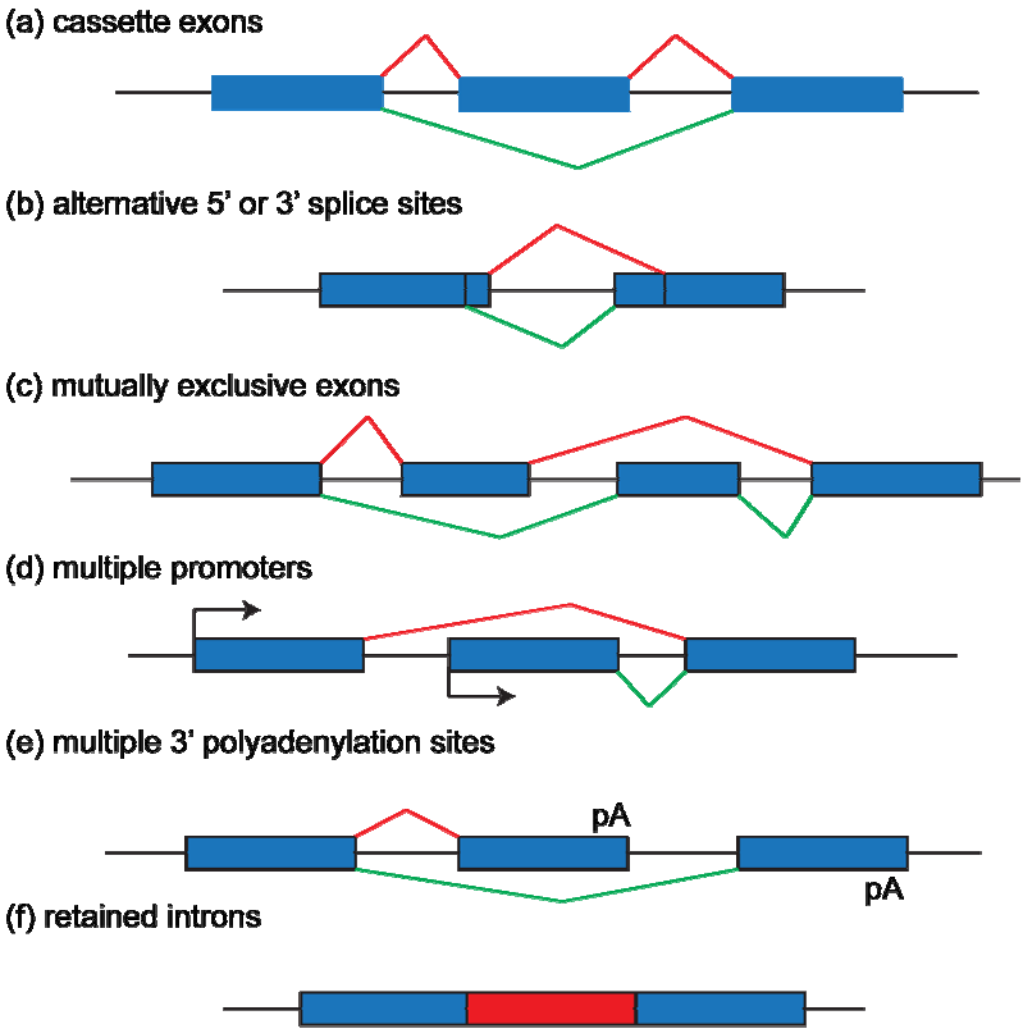


Figure 1.7: Patterns of alternative splicing.
Alternative splicing can generate multiple mRNA isoforms from a single locus through the methods summarised here. Cassette exons (a) may be included or excluded from the mRNA. Alternative 5' or 3' splice sites (b) can lengthen or shorten the exon sequence. Exons can be included in a mutually exclusive manner (c) such that only one exon from many is included in the mature transcript. Alternative splicing at the beginning or end of a transcript can result in use of different promoters (d; indicated by arrows) or polyadenylation sites (e; denoted by pA). Finally, intron retention (f) leads to intronic sequence remaining in the mature mRNA and being translated. These patterns can occur in within the same pre-mRNA leading to a wide repertiore of mature messages from one locus. Figure modified from Black (2003).

Alternative splicing and neurodegeneration

Single base pair substitution mutations that affect AS account for approximately 10% of known human disease-causing mutations, and the majority of these mutations occur within 20bp of a splice site (Krawczak *et al.*, 2007). Mutation of a conserved splice site sequence can result in exon-skipping or use of a cryptic splice site, whereas mutations which affect regulatory sequences can have more subtle effects. The affects of alternative splicing on disease are reviewed in (Caceres and Kornblihtt, 2002). One example of this is mutations within the *tau* gene which are associated with frontotemporal dementia with Parkinsonism linked to chromosome 17 (FTDP-17). Missense, silent or intronic mutations within the *tau* gene are pathogenic due to alterations in the splicing of *tau* exon 10 (D'Souza *et al.*, 1999). In the *tau* gene exon 10 encodes a microtubule-binding domain, therefore aberrant inclusion or exclusion of this exon can have a profound affect on the function of the tau protein. Pathological mutations within the *tau* gene have also been shown to alter exonic splicing enhancer and silencer sequences, and also an intronic splicing silencer.

ADAR and alternative splicing

As described earlier ADAR2 autoedits its own transcript within an intron creating an alternative splice acceptor site which leads to a truncated protein lacking both dsRBDs and the deaminase domain (Rueter et al., 1999). This provides a negative feedback loop to regulate the amount of active ADAR2, and alternative splicing directly correlates to the activity of the ADAR2 enzyme (Maas et al., 2001).

The ADAR2 knockout mice showed an elevated level of incorrectly processed GluR-B pre-mRNA which retained intron 11 (Higuchi et al., 2000). Intriguingly intron 11 contains the ECS region which facilitates editing at the GluR-B Q/R site, leading to the hypothesis that the processes of editing and splicing are linked. The GluR-B R/G editing site occurs within 2bp of the 5' splice donor site of the mutually exclusive FLIP/FLOP exons, providing an excellent system for studying the effect of editing on alternative splicing. In addition to the coding region editing events there are two

"hotspot" regions within the GluR-B pre-mRNA which are edited within intron 11 at the +60 site (hotspot 1) and the +262/263/264 sites (hotspot 2) (Higuchi et al., 1993). Analysis of editing and splicing of these regions of the GluR-B transcript in a cell culture assay revealed that editing at both the Q/R site and intronic hotspot 2 is required for efficient splicing of intron 11 (Schoft *et al.*, 2007). This could account for the observed editing of mature GluR-B transcript which is close to 100% *in vivo*, as unedited transcript would not be processed correctly. The intronic hotspot 2 editing event may alter a splice-repressor signal as mutations which disrupt the structure of this region did not increase splicing efficiency (Schoft et al., 2007). Furthermore a cryptic splice site within exon 11 was discovered which spliced out the Q/R site altogether, and this cryptic splice site was utilised *in vivo* revealing more complex splicing patterns are associated with this region.

In contrast, editing at the R/G site reduced the efficiency of splicing downstream of the editing event, however the decreased splicing efficiency correlated with an increase in splicing fidelity (Schoft et al., 2007). Inhibition of splicing by ADAR2 was also observed with *in vitro* splicing reactions on a GluR-B fragment containing the R/G site, although the inhibition was partially alleviated by addition of RNA helicase A indicating there was a structural requirement for the inhibition (Bratt and Ohman, 2003). Conversely, an "uneditable" construct with a mutated ECS showed skipping of both of the downstream FLIP/FLOP exons, which produced a transcript containing a premature stop codon (Schoft et al., 2007). These results indicate that editing at the R/G site facilitates correct splicing of the downstream FLIP/FLOP exons. A construct containing an inosine residue at the R/G site mimicking constitutive editing also showed inhibition of downstream splicing indicating the effect is due to the presence of inosine rather that ADAR2 binding (Schoft et al., 2007). Therefore editing at the GluR-B R/G site inhibits downstream splicing, and this is probably due to the presence of inosine at the edited site. When analysing the flanking regions the edited R/G site shows greater similarity than the unedited to the consensus binding site for hnRNP A1, indicating this may be the cause for the observed changes in splicing however further work is required to determine whether hnRNP A1 binds in this region. No correlation between editing and choice of the

mutually exclusive FLIP/FLOP exons was observed in adult mouse brain (Schoft et al., 2007).

There are 5 known editing sites in the coding region of the human serotonin (5-HT_{2C}) receptor transcript which all occur within a dsRNA stem-loop formed between exon 5 and an ECS in intron 5 (Burns et al., 1997). This region also undergoes alternative splicing with one canonical and two alternative exon 5 5'splice sites. Examination of editing and splicing in this region of the serotonin receptor transcript revealed a link between increased numbers of sites edited and use of the canonical 5' splice site (Flomen *et al.*, 2004). As the alternative 5' splice sites produce truncated inactive protein, this indicates a link between editing and production of functional serotonin receptor protein. This effect is likely due to disruption of the stem-loop structure which allows access of splicing factors to the canonical 5' splice site, however the effect of editing on changes in splicing regulatory signals cannot be excluded. A study into editing in malignant gliomas which demonstrate hypoediting revealed an increased abundance of transcripts which utilise the alternative 5' splice site in the serotonin $5HT_{2C}$ receptor, and therefore produce truncated inactive protein (Maas et al., 2001).

Characterisation of the human protein tyrosine phosphatase *PTPN6* gene and its role in acute myeloid leukaemia revealed increased expression of an isoform that retained a 251bp intron (Beghini et al., 2000). Sequence analysis of this isoform demonstrated multiple A-to-G mutations were present in the retained intron, one of these occurred 27bp upstream of the 3' splice site and altered the branch point adenosine residue to an inosine residue. Loss of the branch point adenosine led to retention of the intron, which was confirmed with an *in vitro* splicing reaction. Furthermore the retained intron contains an in frame stop codon which would produce a truncated protein lacking the phosphatase domain. Interestingly, levels of the abnormally spliced isoform were significantly higher in patients at the time of diagnosis than when they had entered remission indicating a correlation between *PTPN6* splicing and disease state.

In *Drosophila* editing events were linked with downstream alternative splicing decisions but upstream splicing events were unaffected (Agrawal and Stormo, 2005). This correlation was only demonstrated with one transcript and requires further investigation; however this is the first study into the effects of editing on splicing decisions which occur some distance away from the editing event.

Further work on the edited transcript *resistance to dieldrin* (*rdl*) initially identified in the Hoopengardner screen (Hoopengardner et al., 2003) has revealed the extent to which RNA editing and alternative splicing impact on the function of this transcript. *Rdl* encodes a γ-aminobutyric acid (GABA) receptor subunit, which contains naturally occurring point mutations at amino acid alanine 302 that confer resistance to cyclodiene insecticides (ffrench-Constant *et al.*, 1993). *Rdl* is alternatively spliced at two mutually exclusive exons to produce four different isoforms (ffrench-Constant and Rocheleau, 1993). In addition it can be edited at four different positions which all result in amino acid changes, resulting in 96 potential variant from one genomic locus. However the isoform expression in adult CNS is not random with a 300-fold difference between the least abundant splice-form and the most abundant splice-form (Jones *et al.*, 2009). Analysis of editing isoforms showed that one site (I283V) was edited to ~95% in all isoforms whilst the other three sites showed lower levels. RDL subunits form homo-oligomeric receptors providing an excellent system for evaluating the effect of alternative splicing and RNA editing on the function of the receptor. Injection of transcripts encoding 32 different isoforms into *Xenopus laevis* oocytes revealed that the potency of the GABA receptor varied widely (Jones et al., 2009). The combined editing of three sites had a profound effect on the GABA potency in one splice-form but hardly altered the potency of another splice-form. This indicates that both editing and alternative splicing can modulate the potency of the GABA receptor formed, however these processes could also impact on receptor trafficking and assembly as shown for mammalian GluR-B transcripts (Greger et al., 2003).

There are 10 genes encoding subunits of the nicotinic acetylcholine receptor in *Drosophila melanogaster* and four of these are known to be edited. Together with

alternative splicing editing serves to diversify the receptor composition and alter the physiological properties of the receptor. Editing and alternative splicing changes affect transmembrane regions, ligand-binding domains and at a conserved N-linked glycosylation site, which act in concert to produce a wide range of receptor subunits (Sattelle *et al.*, 2005). Five subunits combine to create each nicotinic acetylcholine receptor resulting in a range of receptors with different properties which are generated from a low number of genes through post transcriptional modification, however not all RNA editing events have been functionally tested therefore the full extent of the effects are unknown.

Comparative genomic analysis of the nicotinic acetylcholine receptor subunit 30D (*nAcRα-30D*) transcript across five different insect orders which span ~300 million years of evolution revealed that editing was present in four of the five insect orders analysed however the sites and frequency varied (Jin *et al.*, 2007). Sites which were edited in some species contained a genomically encoded guanine at the corresponding residue in other species, indicating RNA editing may be an evolutionary intermediate state. Interestingly, several editing sites were discovered in the silkworm *Bombyx mori* which were not present in the other species analysed. Editing at these sites correlated with inclusion of two exons (3a and 3b) which were normally spliced in a mutually exclusive manner in the other species examined. This correlation suggests that editing at sites within exon 4 of the *nAcRα-30D* transcript in *Bombyx mori* leads to inclusion of two exons which are normally spliced in a mutually exclusive manner. However transcripts containing both exons 3a and 3b have been isolated from *Drosophila melanogaster* yet no evidence of editing of the neighbouring exon 4 has been found (Grauso et al., 2002).

These studies indicate that where editing and splicing occur in the same region of a transcript the processes can be linked either through alterations in RNA structure (5-HT_{2C} receptor) or through alteration of splicing regulatory sequences (GluR-B R/G site). However the observed interactions between editing and splicing appear to be transcript specific and cannot be extrapolated to other pre-RNAs. Only one study addresses the issue of long-range effects of editing on alternative splicing (Agrawal

and Stormo, 2005) yet this focuses on one transcript, therefore more work is required to address the effects of editing on alternative splicing.

Coordination of editing and alternative splicing

mRNAs are transcribed *in vivo* by RNA polymerase II (RNAP II), and the processes of splicing, 3' end formation and 5' capping are linked to transcription through association of the factors involved in these processes with the C-terminal domain (CTD) of RNAP II (McCracken *et al.*, 1997a; McCracken *et al.*, 1997b). Therefore investigations into the coordination of splicing and editing have focussed on the C-terminal domain (CTD) of RNAP II. Studies utilising a truncated RNA pol II lacking the CTD region revealed that the CTD is required for efficient editing of the rat *Adar2* transcript at the autoediting site (Laurencikiene *et al.*, 2006). This autoediting event creates a cryptic splice site and is therefore linked to alternative splicing; however absence of the RNA pol II CTD did not affect the efficiency of the splicing reaction. Further investigation of the GluR-B transcript revealed that the RNA pol II CTD was only required for efficient editing at the R/G site when splicing occurred concomitantly indicating a role for RNA pol II CTD in coordinating editing and splicing (Ryman *et al.*, 2007). At the Q/R site which is located 24 nucleotides upstream of a spice donor site the effect was similar in that editing enhanced efficient splicing. These results indicate the CTD of RNA pol II is required to coordinate editing and splicing, possibly through delaying the splicing event until editing has occurred thereby allowing these processes to occur sequentially.

Furthermore, ADARs 1 and 2 have been shown to co-localise with large nuclear ribonucleoprotein (lnRNP) particles which contain numerous splicing factors indicating these RNA processing events may be coordinated (Raitskin *et al.*, 2001).

1.7 Project outline & Aims

The diverse roles attributed to the APOBEC family members that have been characterised to date indicate these enzymes perform crucial roles in the innate

immune system and in the maintenance of genome integrity. Therefore the discovery of a putative novel member of the family whose expression is restricted to the germline warrants further investigation. However, several members of the human A3 sub-family show antiretroviral activity that is independent of deaminase function. Crucial to understanding the function of this novel deaminase is determining its substrate specificity and investigating its expression and subcellular localisation during spermatogenesis. This is the first step to identifying the specific substrates of this putative novel deaminase.

The original aims of this thesis were to express and purify recombinant A4 for *in vitro* biochemical analysis to determine the substrate specificity. At the same time an antibody was generated and affinity purified to analyse the expression of endogenous A4 in mouse testis and sperm. This would also aid in the identification of specific substrates through the expression pattern and the subcellular localisation of the protein. Finally a characterised *E.coli* DNA mutators assay was used to determine whether A4 could deaminate cytosine in DNA as other AID/APOBEC family members can. However due to the inherent insolubility of the recombinant A4 protein and its lack of catalytic activity in the *E.coli* DNA mutators assay, along with several external factors, this project was not continued.

Therefore a new project was initiated with the aim of characterising gene expression changes in *Drosophila melanogaster* that lack the *Adar* gene. These flies show locomotion defects and develop age-dependent neurodegeneration; however the neurodegeneration can be suppressed by expression of a catalytically inactive transgene in the cholinergic neurons. This indicates that ADAR has a deaminase-independent role in the CNS. To characterise this role microarray analysis was performed on 5 day old flies and gene expression changes were verified by real time PCR. In vitro binding and editing experiments were performed to determine whether ADAR interacts directly with these transcripts.

To determine if ADAR binding or activity influence alternative splicing and this was responsible for the altered expression levels observed in *Adar*-null flies a custom AS

array was performed. This identified many genes which show altered splicing in the absence of ADAR. Characterisation of two of these genes demonstrated that aberrant polyadenylation occurs in the absence of ADAR suggesting a novel role for ADAR in 3' end processing of transcripts.

Chapter 2: Materials and Methods

2.1 Materials

Table 2.1. Preparation of Materials

4x Lower Tris 36.34g Tris base 8ml 10% SDS H_20 to 200ml pH to 8.8 with HCl	**4x Upper Tris** 6.06g Tris base 4ml 10% SDS H_20 to 100ml pH to 6.8 with HCl
PBS 137 mM NaCl 2.7 mM KCl 10 mM Na_2HPO_4 2 mM KH_2PO_4	**PBS-T** PBS 0.2% Tween-20
Blocking Buffer for Immunoblotting 5% Milk Powder in PBS-T	**Blocking Buffer for Immunostaining** 1% BSA, 5% serum in PBS-T
Carnoy's Fixative 6 volumes ethanol 3 volume chloroform 1 volume acetic acid	**Laemmli buffer (4X)** 20% glycerol 3% SDS 3% β-mercaptoethanol 10mM Tris-HCl pH6.7-7.0 0.1% Bromophenol blue
LB Agar 10 g NaCl 10 g Bacto-tryptone 5 g Yeast extract 15 g Difco Agar Adjust volume to 1 L with dH_2O	**Luria Broth (LB)** 10 g NaCl 10 g Bacto-tryptone 5 g Yeast extract Adjust volume to 1 L with dH_2O
NuPAGE Running Buffer 104.6 g MOPS 60.6 g Tris Base 10 g SDS 3 g EDTA Adjust volume to 500 ml with dH_2O	**Paraformaldehyde (4%)** 4 g paraformaldehyde in 66ml PBS Heat to 60ºC with stirring (in fumehood) Add 0.5 M NaOH until solution clears Adjust volume to 100 ml with PBS, pH7

TBE (10X)	**TBS**
108 g Tris base 55 g Boric acid 9.3 g EDTA Adjust volume to 1L with dH_2O	20 mM Tris-HCl pH 7.5 137 mM NaCl
Transfer buffer	**Tris-Glycine Running buffer (10x)**
4.04 g TrisBase 14.4 g Glycine 200 ml Methanol Adjust volume to 1L with dH_2O	30g Tris base 144g Glycine H_20 to 1 litre

2.2 Molecular Biology Methods

RNA isolation

RNA was extracted TRIZOL reagent (Invitrogen) according to the manufacturer's protocol. For *Drosophila*, RNA was extracted from 50 mutant or 100 wild type fly heads which had been collected and stored in 500µl TRIZOL at -70 ^{o}C. Briefly, fly heads were homogenised in 500µl TRIZOL and then an additional 500µl TRIZOL was added and samples were left at room temperature for 5 mins. 200µl chloroform was added to the samples and they were shaken vigorously for 15 secs then centrifuged at 12,000 x *g* for 15 mins at 4^{o}C to separate the phases. The upper aqueous phase containing the RNA was transferred to a fresh tube and the RNA was precipitated by addition of 500µl isopropanol. Samples were mixed and incubated at room temperature for 10 mins before centrifuging at 12,000 x *g* for 10 mins at 4^{o}C. RNA pellets were washed with 1ml 75% ethanol and air dried. RNA was resuspended in RNase-free water at room temperature for 10 mins and stored at -70^{o}C. RNA was treated with Turbo DNA-free kit (Ambion) prior to first strand synthesis according to the manufacturer's instructions. Briefly, 1µl Turbo DNase I was added per 10µg RNA, Turbo DNase I 10x buffer was added to a final concentration of 1x, and 1µl RNAse IN+ RNase Inhibitor (Promega) was added. Samples were incubated at 37^{o}C for 30 mins. Turbo DNase I was removed by

addition of 0.1 volume of DNase Inactivation reagent supplied with the kit, samples were mixed and incubated at room temperature for 5 mins. DNase Inactivation reagent was removed by centrifuging samples at 10,000 x *g* for 1.5 mins and RNA was transferred to a fresh tube. RNA was quantified with a Nano Drop ND-1000 Spectrophotometer or when performing microarray analysis samples were analysed with the Agilent Bioanalyser (Agilent). RNA integrity was analysed by gel electrophoresis on a 2% agarose gel in 1x TBE buffer and RNA was visualised with SYBR Green I nucleic acid stain (Invitrogen) diluted 1:10,000 in 1x TBE buffer. Prior to loading onto the gel RNA samples were resuspended in deionised formamide with 1x gel loading buffer (supplied with DNA ladder from Promega). Samples were heated to 65^{o}C for 5 mins to denature the RNA, chilled on ice for 5 mins, and then loaded on the gel.

cDNA synthesis

First strand cDNA synthesis was performed with M-MLV Reverse Transcriptase, RNase H minus, point mutant (Promega) using the manufacturer's protocol. cDNA was synthesised with oligo dT primers (Promega). Briefly 1µg RNA was denatured in the presence of 50ng oligo dT primers at 70^{o}C for 5 mins and chilled on ice. MMLV RT reaction buffer was added to a final concentration of 1x along with 1mM dNTPs and 1µl MMLV RT enzyme. Nuclease-free water was added to a final volume of 25µl. The reaction was incubated at 40^{o}C for 10 mins then at 50^{o}C for 90 mins.

Polymerase Chain Reaction (PCR)

PCR was performed with 1µl DNA or cDNA template (~40ng). Fast start taq (Roche) was used for standard PCR reactions and Phusion polymerase (NEB) was used when high fidelity amplification was required for cloning or analysis of RNA editing frequency. Both enzymes were used as per the manufacturer's instructions. A typical PCR reaction followed these cycling conditions: 95^{o}C for 5 mins denaturation followed by 30 cycles of (95^{o}C- 1 min, 55^{o}C- 30 secs, 72^{o}C- 1 min) and

a final elongation step of 5 mins at 72°C. Time and temperature were varied depending on the primers used and the length of the product. Sequences of the primers used in this thesis are given in Appendix II, Table AII.4.

Real time PCR

Real time PCR was performed with gene-specific primers designed to amplify products of 100-150bp with SYBR Green PCR master mix (Applied Biosystems) on a BioRad CFX96 real time PCR machine. Quantification was performed in Microsoft Excel.

Gel electrophoresis

PCR products were separated by electrophoresis on a 1.5% agarose gel in 0.5X TBE buffer supplemented with ethidium bromide at a final concentration of 0.5µg/ml. Sizes of DNA fragments were estimated with the assistance of either 1 kb DNA ladder (Promega) or 100 bp DNA ladder (Promega). Samples were loaded in an appropriate volume of 6X gel loading buffer supplied with DNA ladder (Promega).

PCR purification by gel extraction

PCR products were electrophoresed on a 1.5% agarose gel and extracted with a Qiagen gel extraction kit and protocol (Qiagen). The manufacturer's instructions were followed and PCR products were eluted in water which had been heated to 50°C to aid elution.

Phenol/chloroform extraction and ethanol precipitation

The sample was adjusted to 100µl volume and an equal volume of phenol/chloroform (Ambion) was added to the sample, followed by centrifugation at 12,000 x *g* for 15 mins. Organic and aqueous components divided into phases and the aqueous, DNA containing, top phase was removed. DNA was precipitated from the aqueous phase by the addition of 3M sodium acetate at 1/10th of the sample

volume, and 2.5 volumes of 100% ethanol. The DNA was centrifuged for 15 mins at 12,000 x *g* at 4ºC. The pellet was washed with 70% ethanol to remove residual salt and left to dry at room temperature. The DNA was re-suspended in an appropriate volume of RNase free water.

Sequencing

Big Dye v3.1 (Applied Biosystems) sequencing kit was used. The manufacturer's instructions were followed and the sequences were run on an ABI Prism 3730 Genetic Analyser (MRC HGU Core Facility). Sequences were analysed using either Sequencher Version 4.0.5 software or Lasergene Seqman software.

pGEM-T easy cloning

Taq polymerases such as Fast start (Roche) frequently add 3' deoxyadenosine to the ends of PCR products in a non-template directed fashion, however high fidelity enzymes such as Phusion polymerase do not. Therefore to A-tail Phusion PCR products the gel purified PCR product was incubated with 1µl 2mM dATP, 1µl Fast Start taq (Roche), and 0.1 volumes of the 10x PCR buffer supplied with the enzyme. Reactions were incubated at 95ºC for 5 mins and then 70ºC for 30 mins. Fast start PCR products were gel purified prior to ligation into the PGEM T-easy vector. PCR products were added to the Promega ligation mixture, according to the manufacturer's instructions and incubated at 16ºC overnight. 5µl of the ligation mixture was transformed into chemically competent *E.coli* XL1 Blue cells.

2.3 Bacterial Methods

Bacterial transformations

5µl of the ligation mixture was incubated on ice for 20 mins with 50µl of chemically competent *E.coli* XL1 Blue cells. The mixture was then heat shocked at 42ºC for 30 secs, before being left on ice for a further 2 mins. 900µl of antibiotic-free LB was added and the mixture incubated at 37ºC with shaking for 1 h. The transformed cells

were spread onto agar plates containing the appropriate antibiotic and incubated overnight at 37ºC. If blue/white selection was employed then X-gal (50ng/ml) and IPTG (500ng/ml) were added to the plates before plating the transformed cells.

Bacterial Cell culture

400ml LB were inoculated with *E.coli* and the appropriate antibiotic (ampicillin at 20mg/ml or kanamycin at 20mg/ml) was added. Cultures were grown in an Innova 4200 incubator shaker (New Brunswick Scientific) set to 220 rpm for >16 h at 37ºC.

Isolation of plasmid DNA

Plasmids were isolated from the positive transformants with QIAGEN mini-prep kit and QIAGEN maxi-prep kit according to the manufacturer's instructions.

2.4 Cell Culture Methods

p53-/- mouse embryonic fibroblast cell culture

p53-null mouse embryonic fibroblasts (MEFs) kindly donated by Jamie Hackett (Richard Meehan's lab) were maintained in DMEM media (Invitrogen) supplemented with 10% fetal calf serum (FCS), 1% penicillin/streptomycin, and 0.5ml β-mercaptoethanol, and grown at 37ºC in the presence of 5% CO_2.

Drosophila S2 cell culture

Drosophila melanogaster S2 cells were kindly donated by Betrand Vernay (Andrew Jackson's lab). *Drosophila* S2 cells were maintained in Schneider's media (Invitrogen) supplemented with 10% FCS and 1% penicillin/streptomycin in a 26^{o}C incubator.

Transfection protocols

p53-/- MEFs were grown to approximately 75% confluence and transfected with a construct encoding GFP-tagged wild type or catalytic site mutant A4 using Lipofectamine-2000 (Invitrogen) according to the manufacturer's protocol.

Drosophila S2 cells were counted and $1x10^7$ cells were transfected with 20µg of plasmid DNA encoding *Drosophila* ADAR-FLIS6 using 100µl Insect GeneJuice reagent (Novagen) in media lacking antibiotics and serum. Cells were incubated with transfection complexes for 4-5 h before it was replaced with fresh media with serum and antibiotics. 24 h after transfection *Drosophila* S2 cells were treated with 0.6mM copper sulphate to induce expression from the metallothionein inducible promoter. *Drosophila* S2 cells were treated with copper sulphate for 24 h prior to harvesting.

Immunofluorescence

For analysis of immunofluorescence transfected cells p53 -/- MEFs were plated on sterile coverslips coated in 1.5% gelatine. GFP-tagged A4 was visualised by fixing cells in a 4% paraformaldehyde solution (4% pFa in PBS with 0.2% Tween-20 (PBS-T)) at room temperature for 15 mins then washing in PBS-T. For visualisation of GFP-tagged A4 coverslips were mounted directly in Vectashield (Vector Labs) with DAPI (1ng/ml). For immunofluorescence with anti-A4 antibodies coverslips were incubated for 1 h with primary antibodies SKC001, SKC002, or SKC003 at either 1:100 or 1:1000 dilution in PBS with 1% BSA and 5% serum. Coverslips were washed three times and then incubated for 1 h with the secondary antibody Texas Red-conjugated anti guinea pig IgG at either 1:1000 or 1:10,000 dilution. Coverslips were washed three times and mounted in Vectashield (Vector labs) with DAPI (1ng/ml).

For immunofluorescence analysis of *Drosophila* S2 cells, transiently transfected *Drosophila* S2 cells in suspension were dropped onto sterile coverslips coated in concanavalin A (0.5mg/ml). Cells were allowed to settle and adhere for 1 h before

fixing in 4% paraformaldehyde at room temperature for 10 mins. FLAG-tagged ADAR protein was visualised with a mouse anti-FLAG antibody (Sigma) at a 1:10,000 dilution. The cells were washed three times in PBS-T and incubated at room temperature with FITC-conjugated anti mouse IgG secondary antibody (Sigma) for 1 h. Coverslips were washed three times in PBS-T and mounted in Vectashield (Vector Labs) with DAPI (1ng/ml).

Immunofluorescence was visualised with an imaging system comprised of a Coolsnap HQ CCD camera (Photometrics Ltd, Tucson, AZ) Zeiss Axioplan II fluorescence microscope with Plan-neofluar objectives, a 100W Hg source (Carl Zeiss, Welwyn Garden City, UK) and Chroma #83000 triple band pass filter set (Chroma Technology Corp., Rockingham, VT) with the excitation filters installed in a motorised filter wheel (Ludl Electronic Products, Hawthorne, NY). Image capture and analysis were performed using in-house scripts written for IPLab Spectrum (Scanalytics Corp, Fairfax, VA).

2.5 Experimental procedures

Plasmid constructs

The A4-FLIS6 coding region was subcloned into the MCS of the pET21b vector via PCR amplification of the yeast expression vector pPICZA-A4-FLIS6 construct (M.Hogg, Master's rotation project) with primers containing either *Nde I* or *Eco RI* restriction sites. The resulting plasmid was sequenced to confirm an intact coding region and termed pET21b-A4-FLIS6. The pET21b-A4-46-to-the-end-FLIS6 vector was created using *Spe I* sites which flank the FLAG-A4-6xHIS insert in the pPICZA-A4-FLIS6 plasmid. A forward primer was designed to amplify the region from nucleotide 136 forward (which corresponds to amino acid 46, numbering from the start methionine) that contained the FLAG epitope coding sequence and a *Spe I* site at its 5' end. The pET21b-A4-FLIS6 vector was digested with *Spe I* to release the insert and the amplified A4-46-to-the-end-FLIS6 insert was ligated into the vector.

The resulting pET21b-A4-46-to-the-end-FLIS6 vector was sequenced to confirm the insert was intact and in the correct orientation.

RNA Immuno-Precipitation (RIP)

Drosophila S2 cells transfected with pRm-ADAR-FLIS6 were harvested by scraping the tissue culture plate and centrifuging at 1000rpm for 4 mins to pellet, and then they were washed once in ice-cold PBS. Cells were lysed in NET buffer (50mM Tris pH7.5, 100mM NaCl, 0.5% Igepal, 1mM DTT) supplemented with 1x complete protease inhibitor (Roche), 0.1U/µl RNase Inhibitor plus (Promega), and 0.1µg/µl yeast tRNA. Extracts were sonicated for 5 mins and centrifuged at 20,000 x *g* for 15 mins at 4^{o}C. Protein concentration was determined by Bradford Assay and 10% input was reserved as an input sample and added to 1ml TRIZOL reagent for RNA extraction. 50µl FLAG M2 agarose beads (Sigma) were pre-incubated with 100µg BSA, 100µg yeast tRNA, and 100µg herring sperm DNA in 1ml NET buffer for 1 h at 4^{o}C with rotation. Beads were washed three times with 1ml NET buffer with centrifugation steps performed at 1000 rpm for 1 min at 4^{o}C. Protein extracts were added to beads and incubated at 4^{o}C with rotation for 2 h. Beads were collected by centrifugation at 1000 rpm for 1 min at 4^{o}C in a table top microfuge. Beads were washed twice for 5 mins in 1ml NET buffer then transferred to a new eppendorf tube and washed for 15 mins at 4^{o}C, then once more for 5 mins. RNA was eluted from beads by addition of 1ml TRIZOL reagent with gentle vortexing. Precipitated RNA was treated with DNase I (Promega) at 37^{o}C for 30 mins before addition of STOP reagent (Promega) and incubation at 65^{o}C for 10 mins. The DNase I treated RNA was then used in RT reactions to generate cDNA. Real time PCR analysis was performed with a standard curve generated from the 10% input sample.

In vitro *editing assay*

In vitro transcribed and purified pre-mRNAs at a concentration of 50ng per reaction were incubated with 50nM purified recombinant ADAR-FLIS6 protein in binding buffer (50mM Tris-HCl pH7.9, 150mM KCl, 5mM EDTA, 20% glycerol, 1mM

DTT, 1x protease inhibitor tablet/50ml, 0.15mg/ml yeast tRNA, and 0.2mg/ml BSA) and RNase IN+ RNase Inhibitor (Promega). Reactions were stopped by the addition of 500µl TRIZOL reagent and RNA was precipitated. cDNA reactions were performed with a mixture of random hexamer and gene-specific primers with control reactions lacking reverse transcriptase to ensure no DNA contamination was present. PCR reactions were performed with high fidelity Phusion polymerase (NEB) with gene-specific primers and products were gel purified and cloned into pGEM T-easy vector. Editing frequency was determined by sequencing of individual clones.

Oxidised RNA immunoprecipitation

1.5µg total RNA purified from fly heads with TRIZOL reagent as described earlier was incubated with 2.5µg anti-8OHG antibody (Abcam) for 30 mins at 4^{o}C. 20µl protein G sepharose beads which had been washed three times in cold PBS were added to each reaction and incubated overnight at 4^{o}C. Beads were pelleted by centrifugation at 1000 rpm for 1 min at 4^{o}C in a table top microfuge. Non-oxidised RNA was precipitated from the supernatant with acid phenol/chloroform and precipitated with 100% ethanol and 1/10th volume 3M NaAc in the presence of 1µl glycogen carrier. Beads were washed three times for 5 mins in 1ml cold PBS with 0.04% Igepal. Oxidised RNA was eluted by addition of 300µl PBS/0.04% Igepal, 30µl 10% SDS, and 300µl acid phenol/chloroform and this was incubated at 37^{o}C for 15 mins with vortexing every 5 mins. RNA was extracted with acid phenol/chloroform and precipitated with 100% ethanol and 1/10th volume 3M NaAc and 1µl glycogen. Following resuspension all the RNA was used in first strand synthesis reactions with a control of 250ng input RNA which was amplified with or without reverse transcriptase enzyme.

In vitro *transcription*

Template DNA encompassing the coding region with introns and UTRs of candidate transcripts were cloned into the PGEM T-easy vector. The vectors were linearised by digestion with *Nco I* for transcripts transcribed with Sp6 polymerase or *Spe I* for

transcripts transcribed with T7 polymerase. 1µg linearized DNA was incubated with T7 or Sp6 polymerase (Ambion) in a 20µl reaction containing 5x buffer, 1µl RNaseIN+ (Promega), 1µl 10mM riboNTPs, 5µg BSA, 1µl synthetic RNA cap. Samples were incubated at 37^{o}C for greater than 2 h with the addition of extra polymerase after 1 h. DNA was digested by addition of 1.5µl turbo DNase I (Ambion), with 40U RNaseIN+ RNAse inhibitor (Promega) and samples were incubated at 37^{o}C for 45 mins. RNA was precipitated with phenol/chloroform isoamyl alcohol (Ambion) in the presence of 1µl glycogen. RNA was purified by gel electrophoresis on a denaturing polyacrylamide gel (4% for longer transcripts 8% for short), and bands were visualised by UV shadowing. Bands were excised with clean scalpel blades and RNA was extracted by incubating the gel slice in RNA extraction buffer (50mM Tris pH7.5, 100nM NaCl, 10mM EDTA and 0.1% SDS) overnight at 4^{o}C on a rotator. RNA integrity was analysed by agarose gel electrophoresis in a loading buffer containing deionised formamide, and RNAs were stored at -70^{o}C in the presence of RNase IN+ RNase inhibitor (Promega).

Radiolabelled substrates were transcribed as above but the riboNTP mixture was adjusted such that rATP, rCTP, and rUTP were included at 10mM and 3µl radioactive rGTP (α^{32}P-rGTP) was included per reaction supplemented with 1µl cold 2mM rGTP. Transcripts were purified by gel electrophoresis on a denaturing polyacrylamide gel and bands were detected by exposing the gel to a film. Bands were excised and RNA was purified as above.

Electrophoretic Mobility Shift Assay

Electrophoretic shift mobility assays were performed by incubating purified ADAR-FLIS6 protein purified from *Pichia pastoris* (Ring *et al.*, 2004) with *in vitro* transcribed radiolabelled *Adar160* RNA. *Adar160* RNA is a fragment of the *Drosophila Adar* transcript which encompasses the self editing site within exon 7 and the ECS. It is edited to approximately 70% *in vitro* (Keegan et al., 2005). The same binding buffer (50mM Tris-HCl pH7.9, 150mM KCl, 5mM EDTA, 20% glycerol, 1mM DTT, 1x protease inhibitor tablet/50ml, 0.15mg/ml yeast tRNA, and

0.2mg/ml BSA) was used for both filter binding reactions and EMSA reactions. Binding reactions were stopped by loading onto a 5% native polyacrylamide gel made with 40% acrylamide: bisacrylamide (29:1), 0.5x TBE, and 1% glycerol to stabilise protein-RNA complexes. Gels were electrophoresed at 4^{o}C for 7-8 h then dried at 80^{o}C for approximately 90 mins. Dried gels were exposed to phosphorimager screens overnight and images were visualised on a FLA2000 laser scanner (IDE). Images were quantified with ImageQuant software (GE Healthcare).

Filter binding analysis of ADAR RNA-binding

Filter binding experiments were performed with *in vitro* transcribed radiolabelled pre-mRNA transcripts and purified ADAR-FLIS6 protein (Ring et al., 2004). Different concentrations of purified ADAR-FLIS6 protein were incubated with α^{32}P-ATP labelled pre-mRNAs in binding buffer (50mM Tris-HCl pH7.9, 150mM KCl, 5mM EDTA, 20% glycerol, 1mM DTT, 1x protease inhibitor tablet/50ml, 0.15mg/ml yeast tRNA, and 0.2mg/ml BSA) for 10 mins before being applied to a 0.45µm nitrocellulose filter (Schleicher & Schuell). A vacuum was applied to the filter and filters were then washed twice with 1ml binding buffer to remove unbound RNAs and air dried. Filters were immersed in 2ml scintillation fluid and counted in a scintillation counter. Bound RNA was calculated using the radioactive count.

RACE

5' and 3' RACE were performed with the Firstchoice RLM RACE kit (Ambion) kit and protocol. 1.5µg total *Drosophila* head RNA was used for each cDNA reaction. RACE products were analysed by gel electrophoresis on a 2% agarose gel in 0.5x TBE. Bands were gel extracted and purified, cloned into PGEM T-easy vector and analysed by sequencing.

SDS-PAGE gel electrophoresis

Protein concentrations were determined by Bradford Assay with BSA standards. Protein samples were mixed with an appropriate amount of 4x laemelli buffer (see Materials section) and boiled at 100^{o}C for 5 mins before cooling to room temperature before loading onto gels. Gels used were either 4-12% Nu-Page Bis-tris pre-cast gels (Invitrogen) which were electrophoresed in Nu-Page gel running buffer (Invitrogen). Alternatively, gels were used which comprised a 4% stacking gel with an 8 or 10% separating gel. 10% separating gels (8.33ml H_2O, 5ml 4x Lower Tris buffer (see Materials section), 6.67ml acrylamide, TEMED 20µl, 10% ammonium persulphate (APS) 100µl) with 4% stacking gel (3.25ml H_2O, 1.25ml Upper Tris buffer (see Materials section), 0.55ml acrylamide, 10µl TEMED, and 20µl 10% APS) were electrophoresed in 1x Tris-Glycine buffer (see Materials section) with 0.1% SDS.

Proteins were electrophoresed through the stacking gels at 15mA and then the current was increased to 25mA for the separating gel. Benchmark pre-stained protein standard (Invitrogen) was included on each gel. Gels were stained with Coomassie brilliant blue stain (50% methanol, 7 % acetic acid, and 0.25g Coomassie brilliant blue R) or Sypro Ruby protein stain (Sigma) according to the manufacturer's protocol.

Silver staining of SDS-PAGE gels

Samples were boiled in SDS buffer (see above) and electrophoresed on a 10% PAGE gel, then fixed in 200mls Fix I (30% ethanol, 10% acetic acid) for 30 mins at room temperature with shaking. Gels were transferred to 200mls Fix II (30% ethanol, 0.5M sodium acetate, 0.2% (w/v) sodium thiosulfate-5 H_2O, 0.125% glutaraldehyde) for 30 mins as before. The gel was then washed three times in distilled water for 10 mins each. The gel was stained for 20 mins in 200mls stain (0.1% (w/v) silver nitrate, 0.01% formaldehyde) and then transferred to developing solution (2.5% sodium carbonate, 0.01% formaldehyde) until stain could be visualized. When stain was visible the gel was transferred to stop solution (1% glycine). Gel images were created by scanning the dried gels.

Western Blot Analysis

Proteins were electrophoresed on SDS-PAGE gels and then transferred to a nitrocellulose membrane (Whatman). Transfer of protein was confirmed by incubating memebranes in Ponceau S solution (Sigma) for 5 mins at room temperature. Excess ponceau S solution was removed by washing in running water and membranes were scanned. Non-specific binding sites were blocked by incubation of the membrane in 5% non-fat dry milk (marvel) in 0.2% PBS-Tween 20 (PBS-T) solution for 30 mins at room temperature. Proteins were detected by incubation of primary antibody dilution in 5% non-fat dry milk in PBS-T for either 1 h at room temperature or overnight at 4^{o}C. Primary antibodies were used at the following dilutions: mouse anti-FLAG M2 antibody (Sigma) 1:3000, guinea pig anti-A4 antibodies 1:100, and mouse anti-SR protein 1H4 antibody (Zymed) 1:500. Membranes were washed three times in PBS-T for 5-10 mins each and then incubated in the appropriate secondary antibody at room temperature for 1 h. Secondary antibodies used were anti-mouse conjugated HRP at 1:10,000 or anti-guinea pig conjugated HRP at 1:5,000. Signals were detected with enhanced chemiluminescence with the Super Signal West Pico detection reagent (Pierce).

2.6 Recombinant protein production and purification

Baculovirus expression of recombinant protein in Sf9 insect cells

Sf9 cells are derived from pupal ovarian tissue from the Fall armyworm (*Spodoptera frugiperda*). They are susceptible to infection by the *Autographa california* multiple nuclear polyhedrosis virus (AcMNPV), and the AcMNPV polyhedrin promoter is used to drive recombinant protein expression in Sf9 cells following infection with the recombinant baculovirus. Wildtype and mutant A4, with and without FLIS6 epitope tags, were cloned into the pENTR vector kindly donated by Tedd Hupp (CRUK) using the Gateway cloning technique (Invitrogen). Plasmids were checked by sequencing and subsequently recombined into the pDEST^{TM}8 vector for use with the Bac-to Bac Baculovirus expression system (Invitrogen). From the pDEST^{TM}8 vector

A4-FLIS6 was recombined into the Baculovirus shuttle vector (bacmid) with the aid of a helper plasmid encoding a transposase, and the bacmid was amplified in *E.coli*. High molecular weight recombinant bacmid DNA was isolated from mini cultures and transfected into Sf9 insect cells.

Sf9 cells obtained from S.Meek (University of Edinburgh) were grown at 26^{o}C with continuous stirring in Sf-900 II serum free media (Gibco) containing 1% penicillin and streptomycin. Plasmids were transfected into Sf9 cells with Cellfectin Transfection Reagent (Invitrogen) following the manufacturer's protocol. Briefly, for each transfection 4µg DNA were mixed with 0.5ml Sf-900 II media minus antibiotics, and separately 8µl Cellfectin was mixed with another 0.5ml Sf-900 II media minus antibiotics. The DNA/media and Cellfectin/media were combined and incubated at room temperature for 15-45 mins. For transfections in 2ml wells, 1 x 10^6 cells were counted and washed in media minus antibiotics. Cell pellets were resuspended in the DNA/Cellfectin/media mix, plated out and incubated at 26^{o}C for 5 h. After 5 h 1ml Sf-900 II media containing antibiotics was added, and cells were incubated for 3 days at 26^{o}C.

Cells were harvested and supernatant containing virus particles was kept to infect future cultures. To prepare a stock of high titre viral supernatant, the supernatant was used to infect 1x10^6 cells, which were grown for 5 days. These cells were harvested and a third round of infection was performed with 2x10^7 cells. The supernatant from the third round of infection was kept and used to infect cells for all further experiments.

Harvested cells were washed once in PBS and the pellet was resuspended in Q200 buffer (50mM Tris-HCl pH8.0, 200mM KCl, 20% glycerol, containing protease inhibitors: 0.7mg/ml pepstatin, 0.4mg/ml leupeptin, 1mM DTT, 1mM PMSF). The cell suspension was sonicated for 3x 10-sec intervals with a microtip sonicator at 15% power and placed on ice for 1 min in between. Samples were then centrifuged at 20,817 x *g* for 15 mins at 4^{o}C. The supernatant was kept for future analysis, and the pellet was resuspended in 100µl Q200 buffer for comparison of soluble and

insoluble protein. Purification with FLAG-M2 agarose and Nickel^{2+} NTA resin were performed as described previously (Ring et al, 2004).

Generation and purification of a polyclonal antibody to A4

Recombinant A4 protein generated for use as an antigen to inoculate guinea pigs and generate a polyclonal antibody was produced in *E.coli* using the pGEX-A4-FLIS6 plasmid (created by A. Ruzov from the pGEX_6P-3 plasmid (Amersham Pharmacia) and the pPICZa-A4-FLIS6 plasmid). The plasmid was transformed into *E.coli* BL21(DE3), which was grown to an $O.D._{600nm}$ of 0.5-0.8 and then induced for 4 h at 30^{o}C by addition of 1mM IPTG. Cells were pelleted at 20,817x *g* for 15 mins at 4^{o}C and lysed for 30 mins at room temperature in 1ml Bugbuster protein extraction reagent (Novagen) and 1µl lysozyme per gram wet weight cells. Lysed cells were centrifuged at 10,000 x g for 30 mins and supernatant was added to 250µl Glutathione Sepharose 4B beads (Amersham Biosciences) that had previously been washed three times for 5 mins each in PBS. The lysed cells were incubated with the beads overnight at 4^{o}C, then washed three times with PBS for 5 mins each. GST tagged FLAG-A4-HIS and GST tagged FLAG-HIS proteins were eluted from beads in 1ml 10mM Glutathione in PBS by end-over end rotation for 1 h at 4^{o}C. This elution step was performed twice, followed by a final elution in 50mM glutathione in PBS. Samples were analysed by western blot with an anti-FLAG antibody. Eluates were run on a 10% polyacrylamide gel to determine purity and a Bradford Assay was performed to determine the protein content. Recombinant protein was used to inoculate 3 guinea pigs (Eurogentec).

Affinity purification of the A4 antibody

For affinity purification of the guinea pig anti-A4 antibodies recombinant GST-A4-FLIS6 protein was generated as described above for immobilization onto a column matrix. To remove antibodies which recognise the epitope tags on the recombinant GST-A4-FLIS6 antigen an additional protein was generated. The pGEX-A4-FLIS6 plasmid was digested with *SpeI* enzyme to remove the A4 coding sequence, creating

the pGEX-FLIS6 plasmid, which encodes a GST-FLIS6 protein. Both the original pGEX-A4-FLIS6 plasmid and the modified pGEX-FLIS6 plasmid were transformed into *E.coli* BL21(DE3) and recombinant proteins were produced as described above. Purification varied from the protocol above as proteins were eluted from beads with 1ml 10mM Glutathione dissolved in Coupling Buffer (0.1M Sodium carbonate, 0.5M Sodium chloride, pH8.3: for the following immobilization of the protein to a matrix) by end-over end rotation for 1 h at 4^{o}C. This elution step was performed twice, followed by a final elution in 50mM glutathione in coupling buffer.

For affinity purification of the guinea pig anti-A4 antibodies two different columns were generated. One column matrix was bound to the GST-A4-FLIS6 protein which was used as antigen to generate the antibodies, and the other column matrix was bound to the GST-FLIS6 protein and used to remove antibodies which recognised the epitope tags. To generate these columns purified proteins were immobilized on a Cyanogen Bromide (CNBr)-activated sepharose matrix (GE Healthcare), as described in the manufacturer's protocol. Briefly, 1g lyophilized CNBr-activated sepharose 4B powder was activated by suspension in 1mM HCl and washed with approximately 200mls 1mM HCl for 15 mins on a sintered glass filter. The protein to be coupled to the matrix was already in suspension in the coupling buffer; this was added to the matrix and incubated overnight at 4^{o}C with end-over-end rotation. The following day the gel matrix was loaded onto a column and washed with 5 column volumes of coupling buffer. Remaining active groups were blocked with 2 h incubation in 0.1M Tris-HCl (pH8.0), and uncoupled proteins were eluted with three cycles of alternating pH (0.1M acetic acid pH4.0 containing 0.5M NaCl, followed by 0.1M Tris-HCl pH8.0 containing 0.5M NaCl). The column matrix was then washed with Q200 buffer and stored at 4^{o}C until required.

For affinity purification of the guinea pig anti-A4 antibodies 5ml serum from an A4 inoculated guinea pig was filtered through a 0.22uM membrane filter and diluted 1:1 in Q200 buffer (50mM Tris-HCl pH8.0, 200mM KCl, 20% glycerol, containing protease inhibitors: 0.7mg/mg pepstatin, 0.4mg/ml leupeptin, 1mM DTT, 1mM PMSF). This was incubated overnight at 4^{o}C with end-over-end rotation with the

GST-FLIS6 matrix to remove antibodies raised to the epitope tags as they would bind to this matrix. The following day the flow through was collected and applied to the GST-A4-FLIS6 matrix to enrich for antibodies which recognise A4 antigens. This was incubated overnight at 4°C with end-over-end rotation and the following day the matrix was loaded into a column and the flow through was re-applied to the column twice to ensure maximal binding. The column was washed three times with 10x column volumes of Q200 buffer. Bound anti-A4 antibody was eluted with 0.1M glycine-HCl pH2.5; 1ml fractions were collected and the pH was immediately neutralised with 50µl 1M Tris-HCl pH9.0. Antibody concentration was determined by absorbance at 280nm in a Nanodrop, and 50% glycerol was added prior to long term storage at –70°C.

2.7 Drosophila methods

Fly maintenance and fly strains

Flies were raised on standard cornmeal-agar medium which was prepared by technicians in the Michael Swann kitchen at Kings building's, University of Edinburgh. Fly stocks were maintained and virgins were collected at 18°C, whilst crosses and amplified stocks were kept at 25°C. Flies were aged in batches of 1-10 flies per vial and they were tipped on to fresh vials every two days. The genotypes of the flies used are listed in Table 2.1.

Fly strain	Description	Genotype
w^{1118}	Control	*w*
$Adar^{5G1}$	Adar deletion	*y,* $Adar^{5G1}$, w^{1}*/FM7* B^{1} g^{4} sc^{8} sn^{X2} v^{Of} w^{a} y^{31d}
Cha-GAL4	Driver line	*w* ; P{Cha-GAL4.7.4}19B P{UAS-GFP.S65T}T2*
UAS-dAdar	Active Adar	*w; ; UAS-Adar 3/4un 1-4a (III)*
UAS-dAdar EA	Inactive Adar	*w; ;UAS-Adar 3/4 EA 6a /TM3 Sb*
UAS-hADAR2	Human ADAR2	*w; UAS-hADAR2 1-4c3 /SM5 Cy;*
UAS-hADAR2 RRM	Human ADAR2 RRM mutant	*w; ; UAS-hADAR 2 RRM /TM3 Sb*

Table 2.1: Fly strains and genotypes.
The strain description and genotypes of flies used in this thesis. $Adar^{5G1}$ flies were initially described in (Palladino et al., 2000b). Where *B* indicates bar eye which is linked to the *FM7* balancer, *w* indicates white eyes, *y* indicates yellow which gives a yellow-brown body colour, *Sb* insictes stubble which gives short bristles, and *Cy* indicates curly which gives curly wings. The *Cha-GAL4* is available from the Bloomington stock centre reference number 6793.

Open Field locomotion assay

Flies were collected when 1 day old by CO_2 anaesthetic and locomotion was analysed at 5 days with no anaesthesia on the intervening days. The assay was performed in the fly room which is maintained at 20°C. Individual flies were placed into a 5mm Petri dish which had been marked into 7 equal areas, one central and 6 peripheral. The dish was tapped and the number of times a fly crossed the line was counted over a 2 min period. Three measurements were recorded for each fly and an average was calculated, and 10 flies of each genotype were assayed. The standard deviation was calculated using the average value for each fly.

Histology techniques

Decapitated fly heads were fixed in Carnoy's fixative (6: 3: 1, ethanol: chloroform: acetic acid) for 4 h at room temperature with rotation. Fixative was removed and heads were dehydrated through 70%, 80%, 90% and two changes of 100% ethanol for 20 mins each at room temperature. The following steps were performed in a Sakura Tissue Tek VIP tissue processor (Sakura) which was pre-programmed. To clear the alcohol, the fly heads were incubated three times in xylene for 5 mins each at room temperature and then once at 60°C. To embed the heads they were incubated in paraffin wax pre-heated to 63°C for 5 mins. This was repeated three times and the fly heads were orientated and mounted onto blocks for cutting. 8μM sections were cut on a Leica microtome with ACCU-edge low profile blades (Sakura). Sections were floated out in a 42°C waterbath and attached to Superfrost plus slides and baked at 50°C overnight.

To remove the embedding wax slides were incubated in xylene three times for 5 mins each. Xylene was removed by incubating slides in three changes of 100% ethanol for 5 mins each. Slides were then re-hydrated through 90%, 70%, 50%, and 30% ethanol for 2 mins each and then rinsed in water. Slides were incubated in freshly filtered haematoxylin for four mins and then rinsed in running water until the water ran clear. Slides were then dipped in twice into acid alcohol (70% ethanol with 2 drops of concentrated hydrochloric acid) and washed in running water. The

slides were dipped twice into lithium carbonate and washed in running water, then incubated in 1% eosin for 5 mins. Slides were washed briefly in running water and then dipped into 100% ethanol. Slides were incubated in fresh 100% ethanol three times for 5 mins each and then taken through three 5 min incubations in xylene prior to mounting in DePeX mounting media (EMS). Sections were imaged with an imaging system comprising a Qimaging Micropublisher 3.3mp cooled colour CCD camera (Qimaging, Burnaby, BC), Zeiss Axioplan II fluorescence microscope with Plan-neofluar or Plan Apochromat objectives (Carl Zeiss, Welwyn Garden City, UK). Image capture and analysis were performed with IPLab Spectrum software (Scanalytics Corp, Fairfax, VA).

2.8 Bioinformatics

Primer design

Primer pairs were designed with the Primer 3 web-based program (http://frodo.wi.mit.edu/primer3/). Primers for real time PCR were designed using the Universal probe library assay design centre (Roche) which is freely available at (https://www.roche-applied-science.com/sis/rtpcr/upl/index.jsp?id=uplct_030000), however universal probes were not used in this thesis. The universal probe library design centre runs *in silico* PCR reactions to analyse primer pairs. Primer sequences are listed in Appendix II, Table AII.4.

Alignments were performed with the Clustalw2 program from the European Bioinformatics Institute (Labarga *et al.*, 2007). The A4 NES was predicted using the freely available NES predictor program at (http://www.cbs.dtu.uk) and the isoelectric point of A4 was estimated with the free program available at (http://www.scripps.edu/~cdputnam/protcalc.html). The freely available miRNA prediction programs TargetScanMouse (Friedman *et al.*, 2009) and miRBD software (Wang, 2008; Wang and El Naqa, 2008).

Statistical tests

Unpaired t-test was performed with the web-based programs available at Graphpad QuikCalcs (http://www.graphpad.com/quickcalcs/index.cfm). An unpaired t-test was used to analyse the locomotion data by comparing the means of two data sets in this case w^{1118} and $Adar^{5G1}$ flies were compared, and the rescue lines $Adar^{5G1}$*:Cha-GAL4>Adar* and $Adar^{5G1}$*:Cha-GAL4>Adar EA* flies.

Affymetrix Microarray Analysis

Microarrays were analysed with the free DNA-Chip Analyzer (dChip) software (Li & Wong, 2001). Affymetrix CEL files were read into the software and analysed using a model based expression profile with perfect match probes only. Arrays were normalised to the array with the median overall intensity. Pattern searches were used to identify significantly altered genes (>1.2 fold change with a p value of $p< 0.05$) which show altered expression in $Adar^{5G1}$ samples and are rescued by expression of active or inactive *Adar* transgenes in the cholinergic neurons.

Alternative Splicing Array Analysis

The alternative splicing (AS) array data was analysed with the Sparrow analysis package written by Marco Blanchette. The analysis was performed in parallel by Marco Blanchette in the USA and at the HGU with the help of Graeme Grimes. The Sparrow package was written and used in the R software environment. Analysis involved scaling and normalisation of the arrays using the Lowess method to remove variation due to dye bias and this was performed with the marray package in R. Genes were then analysed individually using an ANOVA test to determine whether any of the junction probes for a given gene were significantly altered from the other probes for that gene, and this generated a list of genes with altered junctions. To identify the significantly altered junctions a t-test was performed and the Benjamini-Yekutieli correction was applied to correct for multiple testing (Benjamini and Yekutieli, 2001). To determine the expression of the exon junctions, a net expression (NE) value was calculated from the common exonic probes which are

constitutively present, and this was subtracted from the individual junction probe expression data. These calculations highlight differences in exon junction use irrespective of changes in the NE across the transcripts between samples. Subsequently scripts were written in the R environment to display the data graphically and on the UCSC genome browser (http://genome.ucsc.edu/).

Chapter 3: APOBEC4 characterisation

3.1 Introduction

The APOBEC/AID family of deaminases display a wide variety of functions *in vivo*, with different substrate specificities and expression patterns. What has fascinated scientists for a long time is why these potentially powerful mutators exist and how they are regulated. The discovery of a new member of this family that is highly expressed in the testis is intriguing. The germline genome is heavily protected from mutation, therefore why would an active deaminase be found there? Interestingly it has been observed that the point mutation rate is two-fold higher in the male germline than the female germline (Lander et al., 2001), leading us to wonder if APOBEC4 (A4) could be responsible for generating novel point mutations. However increased proliferation in the male germline may be responsible for the increased mutation rate due to replication errors.

A4 was identified by Rogozin and colleagues via bioinformatic searches for the zinc-coordinating motif (H/C)xE….PCxxC that is conserved throughout cytidine deaminase enzymes (Rogozin et al., 2005). The deaminase motif is conserved from deaminases that work on free nucleotides through the ADAR/ADAT family of enzymes that act on adenosine in RNA to the vertebrate-specific AID/APOBEC family that deaminate C-to-U in DNA and RNA. The search revealed an uncharacterized vertebrate protein family with orthologs in many species from humans to *Xenopus tropicalis*, although the A4 protein was not found in any fish databases. The novel A4 protein family showed greatest similarity to A1 (28%), however it is found to be more widespread in the vertebrate lineage indicating it emerged earlier from a common AID/A2 ancestor (Rogozin et al., 2005).

Analysis of the gene structure of A4 indicates it is not likely to be a pseudogene as there is only one copy present and it contains one intron that is efficiently spliced out in the annotated ESTs from this locus. The putative A4 EST was only found in mouse testis although the human database contained 8 testis ESTs, 1 brain EST and 1 EST from uterus (Rogozin et al., 2005). *In situ* hybridisation was performed on mouse testis and confirmed expression of *Apobec4* in the round spermatids and no

signal was detected in mouse ovaries (Diana Best, data not shown). However, to further define the *Apobec4* expression pattern throughout spermatogenesis and to determine whether *Apobec4* was expressed in mature sperm I performed RT-PCR analysis. The preliminary data published on A4 was generated *in silico*, to analyse its predicted deaminase function *in vitro* we decided to generate a recombinant A4 protein.

To characterize the A4 protein *in vitro* I cloned and expressed recombinant A4 in several different expression systems including *E.coli*, the yeast *Pichia pastoris* and a Baculovirus expression system for infection of Sf9 insect cells. However, the recombinant A4 protein was intrinsically insoluble and despite numerous attempts to purify the small fraction of soluble protein obtained I could not isolate sufficient amounts for *in vitro* analysis. After many attempts I concluded that the soluble recombinant A4 produced was not a homogeneous preparation, as I could not purify it under the same conditions.

To investigate the A4 protein localization *in vivo* we generated an antibody to recombinant GST-tagged A4 protein. The GST tag made the protein more soluble and I was able to purify a sufficient quantity to inoculate three guinea pigs. Specifically, we were interested in determining whether the A4 protein is present in testis and mature sperm, as this would help us to elucidate its biological function. In addition, to determine whether A4 could deaminate DNA an *E.coli* DNA mutator assay was performed. This assay has been used to characterise other members of the AID/APOBEC family, and would therefore allow direct comparison of mutator efficiency.

3.2 Results

Bioinformatic Analysis of APOBEC4

Clustal alignment of mouse proteins from the AID/APOBEC family clearly shows the extra four amino acids inserted in the active site of the deaminase domain (Figure 3.1: motif C). These extra bases fall within the zinc-coordinating motif, yet the residues that make contact with the zinc molecule (histidine and cysteine) and the proton donor for the nucleophilic attack on the target residue (glutamic acid) are conserved throughout the family. This suggests the active site may be larger in A4, which we hypothesised could have evolved to accommodate a modified residue such as 5-methylcytosine. A1 and AID have both been shown to be capable of deaminating 5-methylcytosine *in vitro* (Morgan et al., 2004). This indicates that a larger active site may not be necessary for this catalysis, however it may have evolved separately and specifically in A4. A4 may be able to catalyze the deamination of 5-methylcytosine at a faster rate than the other two enzymes.

Interestingly, A4 does not share the conserved SWS motif immediately preceding the PCx(2-6)C motif, instead it contains an NNS sequence (Figure 3.1: motif B). A2 does not share this SWS motif either, but only differs in one residue with SSS at the corresponding position. Strikingly, an A1 mutant (W90S) that contains the A2 sequence SSS behaves similarly to A2 in an *E.coli* DNA mutator assay, showing no mutator activity (Harris et al., 2002). A similar loss of function phenotype was observed with mutations in the core catalytic glutamic acid residue (E63A) and a cysteine residue (C93A) which coordinates the zinc atom in the active site of A1. This indicates these residues play an important role in the deamination reaction, possibly in substrate recognition as they are located so close to the active site. This observation does not imply that A4 is catalytically inactive; these residues may confer different substrate specificity on A4, as is potentially the case for A2. Furthermore, the behaviour of A1 in this *in vitro* assay is not biologically relevant as it has been demonstrated that the *in vivo* substrate of A1 is RNA not DNA.

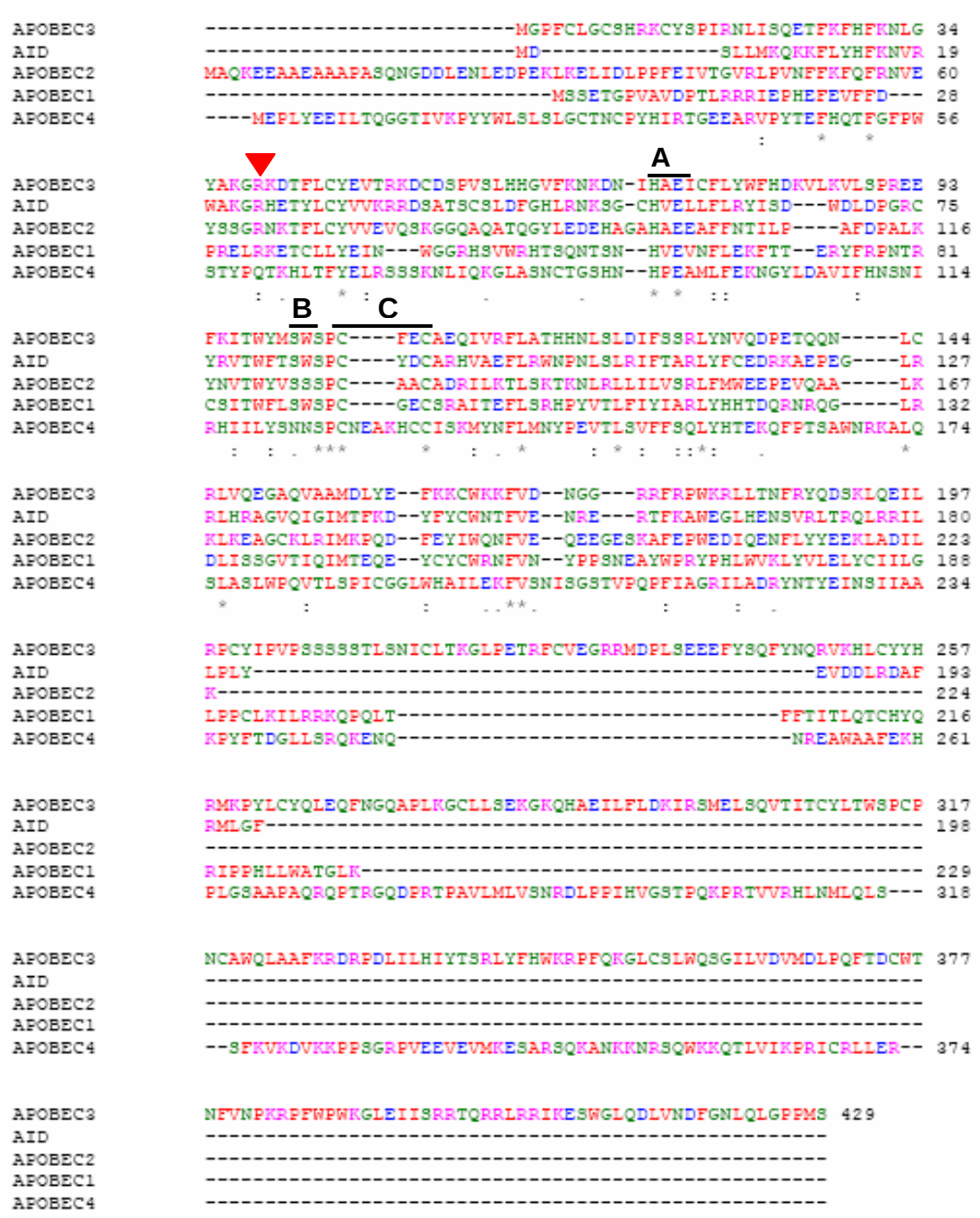


Figure 3.1: Clustal analysis of *Mus musculus* APOBEC/AID proteins.
Alignment of APOBEC 1, 2, 3, 4 and AID protein sequences generated by clustal analysis software from the European Bioinformatics Institute (Labarga *et al.*, 2007). Black bars indicate functional motifs; red triangle indicates a functional conserved arginine residue. Motif A (HxE) which acts as a proton donor during the deamination reaction, motif B (SWS) which may be important for substrate recognition and motif C (PCx(2-6)C) which coordinates the zinc atom in the active site of the deaminase. A4 shows 4 extra amino acids inserted into motif C not seen in other family members. The red triangle indicates an arginine residue involved in allowing substrate access to the active site. Where * indicates complete conservation across all proteins and : indicates an amino acid of similar charge.

Several other amino acids are conserved in the AID/APOBEC family members as distinct from other CDA family members. Intriguingly, A4 seems to share some of these amino acids but diverges at others. Specifically a number of conserved leucine residues downstream of the PCx(2-6)C motif are conserved in A4, whilst arginine and glycine residues share by other family members are not present in A4 (Conticello et al., 2005). As yet no function has been attributed to these residues.

Furthermore, mutational analysis of the AID protein based on predictions from the crystal structure of A2 highlighted the arginine residue found at position 24 in AID and 65 in A2 as being crucial for function (Prochnow et al., 2007) (Figure 3.1 red triangle). The crystal structure showed that this residue is involved in stabilising an "open" hairpin conformation that allows access to the active site. The substitution of a glutamic acid at this residue in the AID protein results in a "closed" loop conformation and deamination activity is lost. In further support of this observation several patients with hyper-IgM-2 syndrome caused by a loss of AID function were found to have a R24W mutation, indicating this site is crucial for AID function (Revy et al., 2000). The arginine is conserved across other APOBEC family members but a glutamine residue is found at the corresponding position in A4.

Human A3C and A3G are also expressed in the testis; therefore differences between A4 and other family member may reflect different functions. Finally, analysis of multiple APOBECs has revealed that many members of this family are under strong positive selection, and as such have evolved at a fast rate. It is entirely likely that differences in A4 reflect further evolution of this gene, although analysis of synonymous and non-synonymous amino acid changes is required to determine this.

Alignment of A4 protein sequences from human, macaque monkey, rat and mouse showed that these proteins were highly conserved at the amino terminus where the deaminase domain is located (Figure 3.2, where red boxes indicate functional motifs within the deaminase domain). This indicates that the function encoded at this region is protected from mutation over time and is of some benefit to the fitness of the organism. However there is considerable divergence at the carboxy terminus

indicating the function of this region is less well protected from mutation or has a less defined function.

Figure 3.2: Alignment of A4 proteins across species.
A4 amino acid sequence from homologues across mammalian species were aligned by clustal analysis software from the European Bioinformatics Institute (Labarga *et al.*, 2007). Species shown are *Rattus norvegicus*, *Mus musculus*, *Homo sapiens*, and *Macaca mulatta*, where functional motifs are indicated by red boxes. There is high conservation across the fist 180 amino acids of the A4 protein, but the conservation is considerably lower at the C-terminal region. Where * indicates complete conservation across all proteins and : indicates an amino acid of similar charge.

Conserved domain searches revealed only the deaminase domain, however a short coiled-coil domain was found at the carboxy terminus between residues 334-354. This is not present in other members of the AID/APOBEC family. BLAST searches with this domain did not reveal any significant homology. Coiled-coil domains are often involved in protein-protein interactions indicating A4 may bind to co-factors or regulatory proteins. Many APOBEC family members contain both nuclear localization signals (NLS) and nuclear export signals (NES) to allow shuttling between the nucleus and cytoplasm. A putative NES was found centred around lysine residue 317 using a NES predictor site **(**http://www.cbs.dtu.dk), however no putative NLS was found. This indicates the A4 protein may enter the nucleus by diffusion as occurs with single-deaminase domain A3 proteins, or it may require a co-factor to be imported into the nucleus. AID and A1 are capable of shuttling between the nucleus and the cytoplasm as they contain both bipartite NLS signals in the N-terminal region and carboxy terminal NES (Bennett *et al.*, 2006; Chester *et al.*, 2003; Ito *et al.*, 2004). Import into the nucleus is dependent on importin α binding, and as the NLS signal occurs within a RNA-binding domain, nuclear localisation does not occur when the APOBEC/AID proteins are bound to RNA. A1 is exported from the nucleus by exportin 1 (Chester *et al.*, 2003). In contrast, human A3G lacks an NLS and is actively maintained in the cytoplasm, even when fused to a strong SV40 NLS sequence (Bennett et al., 2006).

The isoelectric point of the protein was expected to be 9.5 using a charge predictor site (http://www.scripps.edu/~cdputnam/protcalc.html), and the molecular weight was predicted to be 42.7kDa.

Expression Analysis of APOBEC4

Previously *in situ* hybridisation analysis performed by Diana Best confirmed that A4 is expressed in round spermatid cells in the mouse testis. To further clarify the expression of A4 throughout spermatogenesis, RT-PCR was performed on RNA extracted from mouse testis at defined points during development {kindly donated by Howard Cooke} (Maratou *et al.*, 2004). The first wave of spermatogenesis in mice is

synchronous, allowing isolation of RNA from specific cell types by time point analysis. Time points analysed were 7, 11, 15, 18, and 22 days post partum (dpp) and adult mouse testis, which were all taken from C57Bl6 mice. The cell types present in these samples are: A and B spermatogonia, zygotene spermatocytes, early and middle pachytene spermatocytes, late pachytene spermatocytes, and round spermatids respectively. Adult mouse testis includes all cell types as spermatogenesis occurs continuously after birth. Supporting somatic sertoli cells are also present at all stages. The RNA was polyA+ purified and 1ug was used for first strand synthesis. RT-PCR analysis revealed that A4 expression was restricted to late stages of testis development, with a full length transcript not detected until 22 dpp and remaining present in adult testis (Figure 3.3). This is consistent with induction of A4 expression in the round spermatids, although further analysis is required to determine if expression persists in mature sperm. Analysis of the GAPDH transcript revealed the RNA and cDNA were intact, and amplification of samples lacking reverse transcriptase (RT-) indicates there was no DNA contamination.

It has been shown recently that the expression of AID is regulated by miR-155 through a target site in the 3'UTR of the *Aid* transcript (Dorsett et al., 2008; Teng et al., 2008); therefore I investigated the 3'UTR of the *Apobec4* transcript with several miRNA target prediction software programs. The TargetScanMouse software (Friedman *et al.*, 2009) predicted 10 miRNAs could bind to mouse *Apobec4* 3'UTR, but miRBD software (Wang, 2008; Wang and El Naqa, 2008) predicted only 2 miRNAs. Comparing both lists revealed one miRNA in common; miR-741 showed an exact match at the seed region to nucleotides 1518-1524. Interestingly this region of *Apobec4* is conserved in humans, chimpanzees and rhesus monkeys. Further investigation of the expression pattern of miR-741 revealed that it was recently identified as one of a novel group of miRNAs transcribed from the X chromosome during spermatogenesis (Watanabe *et al.*, 2006). miR-741 expression is testis-specific and increases during spermatogenesis, being undetectable at 8 dpp then present at both 15dpp and in adult testis, similar to the expression profile of *Apobec4*. This indicates that if miR-741 is regulating A4 expression it may be acting to block translation of the *Apobec4* transcript as both miR-741 and *Apobec4* mRNA are

expressed at the same stage of spermatogeneis. However further investigation is required to determine if both miRNA-741 and *Apobec4* are expressed in the same cells and whether miRNA-741 binds to and regulates *Apobec4* transcript levels in a similar manner to which miR-155 regulates AID.

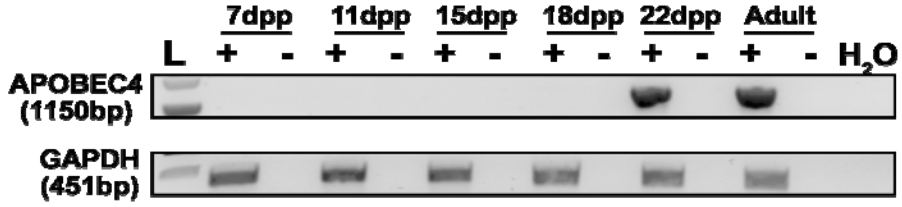


Figure 3.3: APOBEC4 expression analysis during spermatogenesis
APOBEC4 specific primers were used to analyse expression of APOBEC4 during spermatogenesis. Oligo (dT) primed cDNA was synthesised from 1ug total RNA extracted from mouse testis at 7, 11, 15, 18, and 22 days post partum (dpp) and adult. RT positive (+) and RT minus (-) comtrols were included to ensure no DNA contamination was present. A standard PCR program with 55°C annealing temperature and 30 or 25 cycles was used for APOBEC4 and GAPDH repectively. L indicates a DNA size ladder and H_2O indicates a water control. PCR product sizes are indicated. APOBEC4 expression is not detected at early stages of spermatogenesis, but is clearly present at 22 days post fertilization and in adult mouse testis. The GAPDH control is present at all stages indicating the RNA integrity is good.

Expression of recombinant tagged APOBEC4 in E.coli

The coding region of A4 was cloned into the *E. coli* expression vector pET21b with an N-terminal FLAG tag and a C-terminal 6x histidine tag, referred to as a FLIS6 tag hereafter. This combination of two tags has been used in the O'Connell lab to successfully purify recombinant ADARs from the yeast *Pichia pastoris* (Ring *et al.*, 2004). Following transformation into *E.coli* several colonies were grown up in mini-cultures and screened for expression of A4-FLIS6 following induction with 1mM IPTG. Cultures were centrifuged and total extract was boiled in laemelli buffer before loading onto an SDS-PAGE gel. Recombinant A4-FLIS6 was detected using an antibody to the FLAG tag (Figure 3.4). A band of approximately 45kDa is seen clearly in both the coomassie stained gel and the immunoblot, which corresponds to recombinant A4-FLIS6 indicating robust induction and expression of the protein. The predicted molecular weight of A4 is 42.7kDa however the epitope tags cause a

small increase in the molecular weight to 44.5kDa. Expression of recombinant protein was absent in samples before induction indicating robust control in this expression system. Following induction with IPTG recombinant A4 can be seen on both coomassie gel and western blot however expression varied a little between independent cultures. A band of similar molecular weight (~45kDa) is present in uninduced samples, indicating a single or multiple endogenous *E.coli* proteins of similar molecular weight are present.

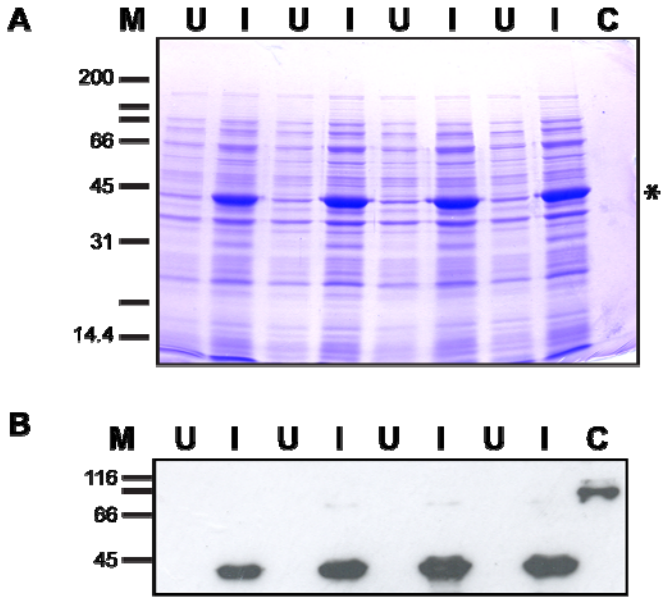


Figure 3.4: APOBEC4 expression in *E.coli*

BL21(DE3) *E.coli* were transformed with pET21b-A4-FLIS6 construct and 4 ampicillin resistant colonies were grown overnight in LB media. 100ul of this culture was used to innoculate fresh LB media (1.5ml). Cultures were grown for 2 hours and an uninduced (U) aliquot was taken. 100mM IPTG was added to induce expression of A4 protein and cultures were incubated for a further 2 hours and an induced (I) aliquot was taken. Aliquots analysed are total extract boiled in Laemmli buffer. Samples were analysed by coomassie gel (A) and immunoblot (B) using an anti-FLAG antibody. A FLAG-positive band of approximately 45 kDa is present in all the induced samples and absent from all uninduced samples as seen by immunoblot analysis. A corresponding band is clearly upregulated on the coomassie gel (asterisk), however a band of the same molecular weight is also present in the uninduced samples. M indicates the molecular weight marker (kDa) and C is a positive control FLAG-tagged protein for the FLAG antibody.

One highly expressing starter culture was used to inoculate a larger 500ml culture. Following induction with IPTG for 2 hours the bacterial cells were harvested by centrifugation. Cells were broken by passing through a French Press three times and the resulting extract was centrifuged to separate soluble from insoluble material. Aliquots were analysed by coomassie gel (Figure 3.5) which indicated the majority of the overexpressed protein was present in the insoluble fraction.

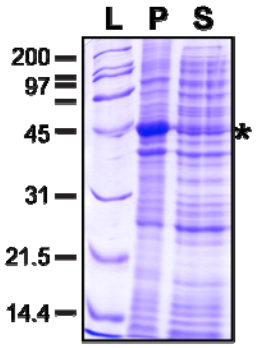


Figure 3.5: Recombinant APOBEC4 expressed in *E.coli* is insoluble.
Recombinant APOBEC4-FLIS6 expressed in *E.coli* following induction with IPTG. For purification *E.coli* cells were broken with a French press and the extract was centrifuged to separate insoluble material. Samples from the pellet (P) and supernatant (S) were analysed by Coomassie gel, which revealed the majority of the recombinant APOBEC4 remained in the insoluble fraction (indicated by asterisk). L indicates molecular weight ladder, size in kDa is listed on the left.

The induction and expression of recombinant A4 indicated sufficient amounts of protein were produced however the majority of the protein was insoluble. It is widely noted that proteins over-expressed in *E.coli* can accumulate in aggregates inside insoluble inclusion bodies (For a review, see (Marston, 1986)). Specific protocols have been developed to purify first the inclusion bodies, and secondly the protein from within it. The protocol I followed was modified from Sambrook & Russell, 2001, and involved two digestion steps to release the inclusion bodies and sequential incubations with detergent and urea to release the protein from the inclusion bodies (See Methods section for detailed description). An *E.coli* culture that had been transformed with the empty backbone vector pET21b was used as a negative control and induced with the same conditions as the A4-FLIS6 culture. For analysis, aliquots of supernatant were taken after isolation of inclusion bodies, after treatment with Triton X-100, and after incubation with alkaline monopotassium

phosphate and urea (S1, S2, and S3 respectively, panel A, Figure 3.6). Coomassie gel analysis revealed that the majority of the recombinant A4 remained in the insoluble pellet despite all these treatments. A band of similar molecular weight to A4 is present in the S1 and S2 fractions, but this band also is present in the control purification indicating it is an *E.coli* protein. The remaining insoluble pellet containing A4 was divided subjected to a longer incubation of 2 hours with buffer containing varying concentrations of urea (2M, 4M, 6M, or 8M, panel B, Figure 3.6) however A4 clearly remained in the insoluble pellet fraction. Very few of the insoluble *E.coli* proteins were solubilzed by 2M urea, as shown by the lack of protein in lane 2 (Figure 3.6, panel A). However, with progressively higher concentrations of urea many of the insoluble *E.coli* proteins initially found in the pellet fraction showed a shift to the supernatant, indicating the conditions used were able to solubilize some of the insoluble proteins (Figure 3.6, compare 2M-S fraction with 8M-S fraction).

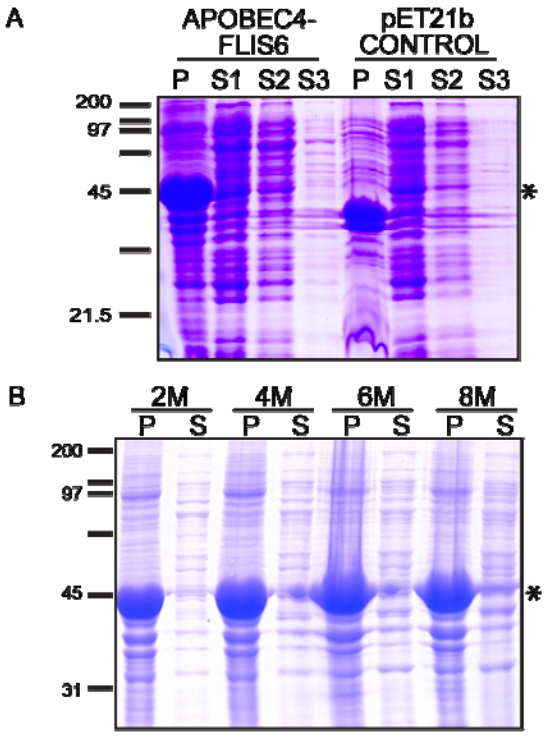


Figure 3.6: UREA solubilization of recombinant APOBEC4
Protein was extracted from *E.coli* expressing A4-FLIS6 using a protocol designed to purify proteins from insoulble inclusion bodies. (A) Following isolation of inclusion bodies several methods were used to try and solubilize recombinant A4-FLIS6, where supernatant 1 (S1) is the starting supernatant, S2 is the supernatant following incubation with a solution containing Triton X-100, S3 is supernatant following incubation in alkaline monopotassium phosphate and urea, and P is the insoulble material that remained following these treatments. A4-FLIS6 seen as a strong band of ~45kDa, remained in the insoluble pellet fraction, a band of similar molecular weight in the supernatant is also present in the control purification. A control extract from *E.coli* transformed with the empty pET21b evctor was purified in parallel. Following this the remaming pellet was divided and incubated for a longer time in different concentrations (2M, 4M, 6M, and 8M) of urea to try and solubilaze the protein (B). A4-FLIS6 remains in the insoluble pellet (P) fraction at all urea strengths tested. In contrast, many of the native insoluble *E.coli* proteins seen in the pellets of the control containing the empty pET21b vector are solubilised at 8M UREA as these proteins are seen in the soluble (S) fraction.

At the time of these experiments a paper was published reporting the crystal structure of human A2 protein produced in *E.coli*, however instead of using the full length protein the authors used a fragment lacking the first 40 amino-acids (Prochnow et al., 2007). At the time there were no other crystal structures published for members of the AID/APOBEC family due to problems with purification. Believing this modification might make recombinant A4 more soluble I replicated this truncation. Analysis of human A2 predicted secondary structure revealed that there was a break between two predicted loop conformations between amino-acids 40 and 41.

Therefore I searched A4 secondary structure predictions to determine where would be best to cleave the protein with minimal disruption. A natural break between two short β-sheets occurred between amino acids 45 and 46, therefore I created a fragment consisting of A4-46-end-FLIS6 by PCR and cloned this into the pET21b expression vector. Following induction with IPTG for 2 hours cells were lysed and purified.

For A4-46-end-FLIS6 purification the 6-histidine epitope tag was utilised by applying the extract to a Ni^{2+} NTA superflow resin (Qiagen), which binds 6-histidine epitopes with a binding capacity between 5-20mg/ml. Following this purification the protein was eluted into one fraction as this method has been successful in purifying recombinant ADAR proteins, and it has been demonstrated that immidazole does not affect binding to anti-FLAG affinity resin (Ring et al., 2004). This was then applied to an anti-FLAG M2 affinity resin (Sigma) which has a binding affinity of >0.6mg/ml. Aliquots were taken for analysis by coomassie gel and immunoblot (Figure 3.7). Recombinant truncated A4 can clearly be seen as a band of ~40kDa in the insoluble pellet fraction (P) on the coomassie gel, and as an intense smear on the immunoblot due to the presence of large amounts of recombinant protein. The immunoblot indicates a small amount of soluble protein is present and successfully purified using the 6-histidine tag as a band is detected in the Ni^{2+} NTA eluate/FLAG load (E/L) fraction. However, the following purification over a FLAG affinity resin was not successful as the protein is present in the flow through (FT) indicating it did not bind to the FLAG resin. Unfortunately the purified protein eluted from the Ni^{2+} NTA resin is not detectable by coomassie staining, indicating the amount is below the 50ng detection limit of the coomassie stain.

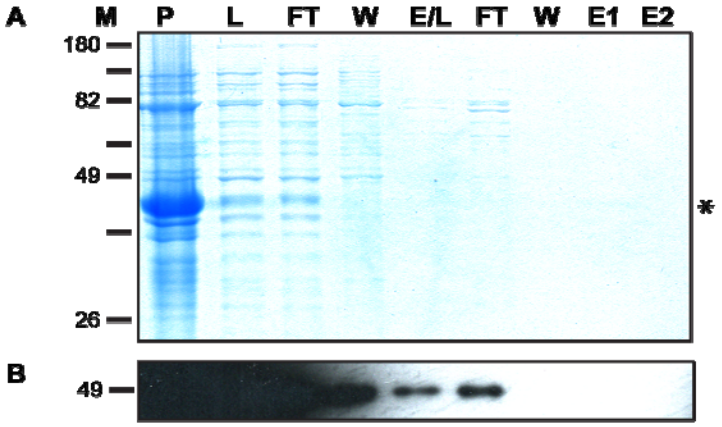


Figure 3.7: Purification of a truncated APOBEC4 protein
A truncated APOBEC4 protein lacking 45 amino acids from the amino terminal but retaining the FLAG and HIS epitope tags was cloned into the pET21b *E.coli* expression vector, expressed and purified. A4 46-END-FLIS6 was purified over a Ni^{2+} NTA column followed by a FLAG affinity column and aliquots were analysed by coomassie stain (A), however the majority of the protein remained in the insoluble pellet (P) fraction. Immunoblot analysis of aliquots taken during the purification process detected with an anti-FLAG antibody (B) confirmed that the majority of the protein was in the insoluble fraction which appears as an intense smear due to the high affinity of the anti-FLAG antibody. Recombinant APOBEC4 protein does not appear to bind to the FLAG resin as a band can be detected in the flow through (FT) from this column. Where P = insoluble pellet, L = load, FT = flow through, W = wash, E/L = elute from the first column which was then loaded onto the second column, and E1 and E2 = eluates 1 and 2 from the final column. M indicates molecular weight marker with size listed on the left (kDa).

In summary, the purification using the Ni^{2+} NTA resin was successful; however as the majority of the recombinant full length or truncated APOBEC4 remained insoluble this system was not sufficient for generating enough protein to biochemically characterize *in vitro*. Amount of IPTG inducer, and the time and temperature of induction were all varied in an attempt to produce less protein, which may have been more soluble but these approaches were not successful so I decided to try a different expression system.

Baculovirus-driven expression of recombinant tagged APOBEC4 in insect cells

Expression of recombinant A4 in *E.coli* resulted in large amounts of insoluble protein therefore a different expression system was required. Baculovirus-driven expression of recombinant proteins in Sf9 insect cells is an efficient method of

producing recombinant protein with many eukaryotic post-translational modifications and can be scaled up easily. For future studies a catalytic site mutant of A4 was generated by Jim Brindle who performed site-directed mutagenesis to change the glutamic acid to alanine (EA), which is the proton donor for the deamination reaction. Analogous mutations have been made in many members of the cytidine deaminase family and shown to abolish deaminase function, such as A1 and ADARs (Harris et al., 2002). Both wild type and catalytically inactive A4-FLIS6 were cloned into a Bac-to-Bac Baculovirus expression vector using Gateway™ cloning technology (Invitrogen), which abrogates the need for PCR therefore reducing the risk of errors in the final vector. The destination vector contained a baculovirus-specific polyhedrin promoter directing expression of wild type A4-FLIS6 or of putative catalytic-site mutant A4 EA-FLIS6. The Bac-to-Bac Baculovirus Expression system (Invitrogen) utilises a high molecular weight baculovirus shuttle vector (bacmid) which can be replicated in *E.coli*. With the aid of a helper plasmid containing a transposase gene, recombinant tagged A4 was transposed from a Gateway pFastBac vector into a recombinant bacmid. Following screening of colonies for presence of the correct insert, high molecular weight bacmid DNA was isolated and transfected into Sf9 insect cells using Cellfectin (Invitrogen). Baculovirus expression of recombinant A4 was performed in the lab of Ted Hupp at the Edinburgh Cancer Research Centre with the help of Sarah Meek. After incubation for 72 hours cells were harvested, cells were checked for expression of recombinant A4 and supernatant containing recombinant virus was kept for amplification and future rounds of infection.

Initial screening of the soluble fraction of cellular extract for the presence of a protein of approximately 45 kDa with an anti-FLAG antibody revealed the Baculovirus Expression system was able to produce soluble recombinant A4-FLIS6 protein (Figure 3.8 A). Approximately 30-40% of the amount in the total extract (T) was present in the soluble supernatant fraction (S), however the majority of the protein remained in the insoluble pellet fraction (P), but this fraction also showed considerable non-specific staining with the anti-FLAG antibody making quantification difficult. To optimise the soluble recombinant A4-FLIS6 production I

performed a time course experiment, analysing samples 24, 48 and 72-hours post-infection (Figure 3.8 B). No soluble A4-FLIS6 was detected 24 hours after infection as analysed by western blot with an anti-FLAG antibody. A band of approximately 45kDa was detected after 48 hours and increased in intensity by 72 hours; however at 72 hours a lower molecular weight band was also detected indicating degraded protein was present. As the degraded protein was detected with an anti-FLAG antibody it must have retained the FLAG epitope, and would probably be purified along with full-length protein if this epitope was used for purification. Therefore 48 hours post-infection was chosen as the best time to harvest cells for future experiments. Further optimisation to determine the optimal infection conditions for soluble A4-FLIS6 production involved infecting $1x10^6$ cells with increasing amounts of recombinant baculovirus (0, 5, 10 15 and 20µl) and analysing protein expression by immunoblot (Figure 3.8 C). The amount of soluble recombinant A4-FLIS6 produced was low with 5µl inoculum and increased when 10µl inoculum was used, however it then decreased with amounts greater than 10µl indicating that viral load was not a limiting factor in the protein production. This also indicates that higher amounts of virus are detrimental to protein expression. This could be a toxic effect of the viral load on the cells or inhibition of protein production, but from previous experiments it is likely due to an increase in insoluble protein, which would reduce soluble protein through aggregation. The ratio of 10ul inoculum per $1x10^6$ cells was used for future experiments. Optimisation was performed with wild type A4-FLIS6, however the catalytic site mutant A4-FLIS6 was generated in parallel.

The baculovirus expression system makes scaling up cultures relatively easy as the virus can be amplified easily and the number of cells infected increased. Therefore, despite the majority of the recombinant protein remaining insoluble I focused on optimising the purification method for the soluble fraction.

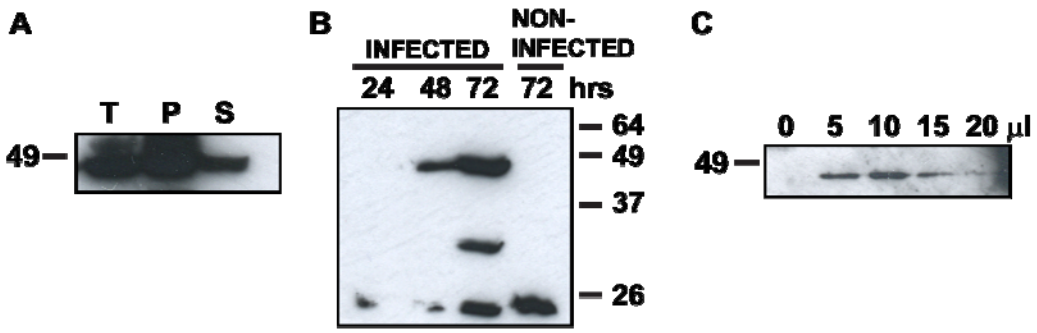


Figure 3.8: Baculovirus-driven expression of recombinant APOBEC4 in Sf9 insect cells.

Supernatant containing baculovirus particles from P1 cultures that have been engineered to express recombinant tagged A4-FLIS6 was used to infect Sf9 insect cells. 20ul of extract was loaded directly for total extract (T), and 20ul was sonicated and centrifuged to obtain an insoluble pellet (P) and soluble supernatant (S) fraction to be analysed by immunoblot (A). A band of approximately 45kDa was present in all three samples but the majority of the recombinant A4-FLIS6 was in the insoluble pellet fraction. However, approximately 30-40% of the amount in the total extract was present in the soluble supernatant fraction (estimated from blot). (B) For further analysis supernatant samples were prepared at 24, 48 and 72 hours post-infection and analysed by immunoblot to determine optimal soluble protein expression. No soluble protein was detected after 24 hours, but by 48 hours a band of approximately 45kDa corresponding to recombinant A4-FLIS6 was observed. By 72 hours a stronger band of 45kDa was observed, however a lower molecular weight band was also present indicating degradation had occurred. An extra band of approximately 20kDa was present in both infected and non-infected samples after 72 hours indicating this is a non-specific protein. (C) To determine the optimal infection conditions for protein production varying amounts (0, 5, 10, 15, and 20µl) of supernatant containing recombinant A4 Baculovirus were used to innoculate $1x10^6$ cells, supernatant was harvested and analysed by immunoblot. A low level of recombinant protein is detected with 5µl innoculum, but higher levels are detected with 10µl. Innoculum volumes greater than 10µl appear to be detrimental to protein expression as soluble A4 levels decrease. Recombinant A4 was detected with an anti-FLAG antibody (Sigma). Molecular weight markers are indicated in kDa.

Initial attempts to purify the soluble fraction of recombinant A4-FLIS6 from Sf9 cells focused on utilising the epitope tags. $1x10^7$ Sf9 cells were inoculated and maintained in identical conditions. They were purified over either anti-FLAG affinity matrix or Ni^{2+} NTA resin to assess which epitope tag performed best. Fractions were analysed by immunoblot and PAGE gels which were silver-stained (Figure 3.9). The results were disappointing as neither epitope tag was able to bind the total amount of soluble A4-FLIS6 in the samples, and in both cases showed protein was lost during the washes. However, it was unclear from these results whether the presence of other proteins in these samples was affecting the ability of recombinant A4-FLIS6 to bind to the matrices. Direct comparison of both

purifications showed the anti-FLAG affinity column performed best, although the silver stained gel revealed the eluate contained many impurities.

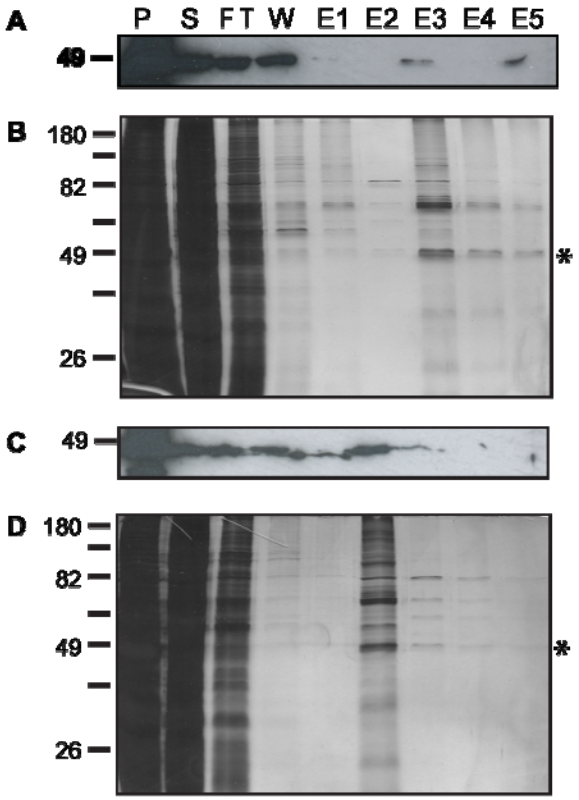


Figure 3.9: Comparison of epitope tags for purification of APOBEC4
(A and B) Soluble recombinant APOBEC4 was isolated from Sf9 cells and purified over either a Ni-NTA column or (C and D) an anti-FLAG affinity column. Fractions were analysed by ummunoblot with an anti-FLAG antibody (A and C) or by silver-staining of SDS-PAGE gels (B and D). Although recombinant A4 bound to the Ni^{2+}-NTA column most protein was lost during the washes, with approximately <5% of the soluble protein present in elution fraction E3, although this fraction is not pure. Recombinant APOBEC4 bound to the anti-FLAG affinity column, and despite some protein being lost during washes aproximately 20% of the protein was present in elution fraction E2. This fraction required further purification to remove contaminating proteins observed on the silver-stained gel. 1ul pellet, supernatant and flow through was loaded, 10ul of other fractions were loaded. Molecular weight is shown in kDa.

Ni^{2+} NTA resin can be used to purify proteins from a crude lysate as it has a high binding capacity (5-20mg/ml). However FLAG M2 affinity matrix has a low binding capacity (0.6mg/ml) therefore it is sometimes necessary to first purify protein over another matrix. As the starting samples were total protein extracts I

thought that purifying these further using other matrices to reduce the amount of protein loaded onto the epitope columns might improve the binding of the epitope tags to the affinity matrices. This should also result in cleaner elution fractions after purification with epitope tags. To achieve this I took the flow through fractions from these columns and tested them on several different matrices, summarised in Figure 3.10. The flow through from the Ni-NTA column in Figure 3.9 was further purified over Affi-Gel Blue Gel matrix (Bio Rad) as this is often used to purify proteins that bind nucleic acids. However, recombinant A4-FLIS6 did not bind to Affi-Gel Blue Gel matrix as a band is detected in the flow through.

The flow through fraction from the anti-FLAG affinity column in Figure 3.9 was loaded onto an anionic exchange resin DEAE sepharose fast flow (Pharmacia Biotech), as the predicted isoelectric point of recombinant A4-FLIS6 is 9.5. All the recombinant A4-FLIS6 bound to the DEAE sepharose column and eluted with 500mM potassium chloride in elution fractions E4 and E5. Further analysis of this purification by silver-stained SDS-PAGE revealed contaminants were still present in these elution fractions; however the total amount of protein present was substantially less than in the load fraction (Compare Figure 3.10 panel C lane marked L with lanes marked E4 and E5). As I had successfully reduced the total protein without loss of A4-FLIS6 I then pooled the DEAE sepharose elution fractions E4 and E5 and loaded them onto an anti-FLAG affinity column. Unfortunately none of the recombinant A4-FLIS6 bound to the affinity column as demonstrated by the presence of a band in the flow through fraction. This indicates that irrespective of the total protein content of the sample there is a limited amount (approximately 20% estimated from Figure 3.9) of recombinant A4-FLIS6 that is able to bind to the anti-FLAG affinity column. The A4-FLIS6 in the initial flow through was unable to bind to the FLAG M2 resin and this was not due to over loading the column. This could be due to several reasons; firstly the epitope tag may not be exposed; and secondly, the epitope tag may be concealed due to multimerisation of the protein, members of the APOBEC/AID family are thought to function as homodimers or homotetramers.

In a final attempt to purify these fractions further the flow through from this anti-FLAG affinity column was loaded onto a poly dI:dC column in the hope that

recombinant A4-FLIS6 might show non-specific binding to RNA. But A4-FLIS6 failed to bind strongly to poly dI:dC matrix and what was bound was lost during the washing steps.

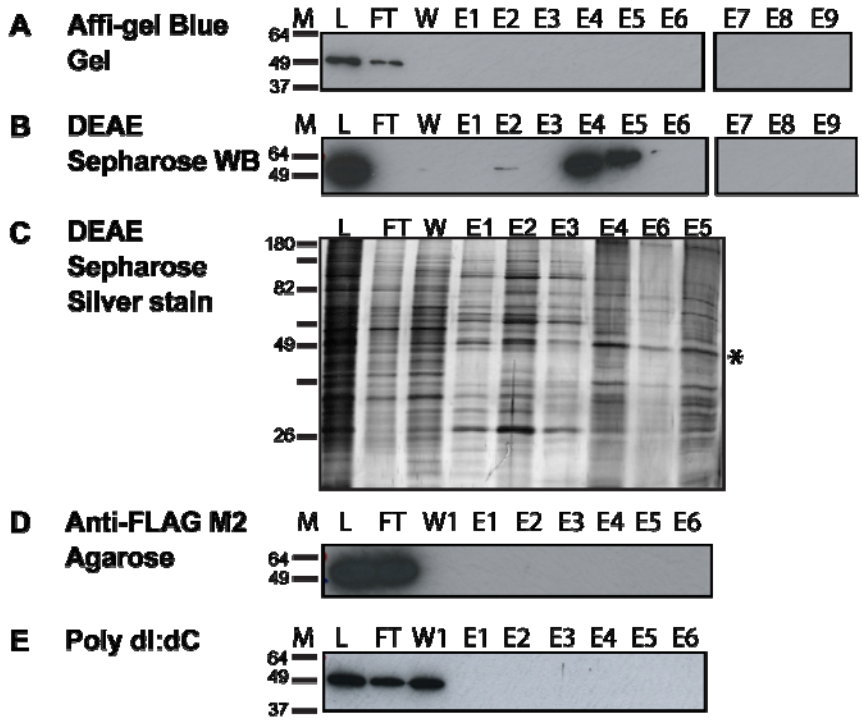


Figure 3.10: Purification of recombinant APOBEC4 with different matrices

Recombinant APOBEC4 was expressed in Sf9 insect cells following *Baculovirus* infection and purified by ion exchange and affinity chromatography, aliquots were taken during purification and analysed by immunoblot with an anti-FLAG antibody. (A) Recombinant APOBEC4 did not bind to Affi-Gel blue matrix as shown by the presence of a band in the flow through (FT). (B) A promising result was obtained with a DEAE sepharose matrix as the protein was bound and eluted in fractions 4 and 5. (C) However these fractions were not pure when analysed by silver stain. (D) Therefore eluates 4 and 5 from the DEAE sepharose column were pooled purified further over the Anti-FLAG Agarose column. Unfortunately, the partially purified protein did not bind to the FLAG resin, as shown by the presence of a band in the flow through. (E) The same partially purified fraction (FLAG flow through) was then applied to a poly dI:dC matix however the bound protein was lost in the wash fraction. L= 1ul load, FT= 1ul flow through, W= 1ul wash, and E1-9= 10ul eluates 1-9. M indicates molecular weight in kDa. Note: DEAE eluates E5 and E6 were accidentally loaded in reverse order on the silver stained gel.

To repeat the most promising result that was obtained with DEAE Sepharose fast flow resin, fresh inoculations were performed, this time the catalytic site mutant A4

EA-FLIS6 was included. $2x10^7$ Sf9 cells were inoculated with 200ul recombinant Baculovirus as determined in the optimisation process performed on WT A4-FLIS6 (Figure 3.8). Cell extracts were analysed by Bradford assay and total protein content was calculated to be 9mg for WT A4-FLIS6 and 10.5mg A4 EA-FLIS6 in 1ml volume. As the binding capacity of DEAE Sepharose is ~100mg/ml therefore the total extract was loaded onto a 1ml DEAE Sepharose resin. The matrix was washed in Q200 buffer and protein was eluted with increasing concentrations of potassium chloride (250mM, 500mM and 750mM). Aliquots were taken and analysed by immunoblot using an anti-FLAG antibody (Figure 3.11). Immunoblot analysis of WT A4-FLIS6 showed that in contrast to earlier results recombinant A4 did not bind to the DEAE Sepharose matrix with high affinity. Only a small fraction of the recombinant A4-FLIS6 was retained on the column and eluted as before with 500mM potassium chloride, the majority of the protein remained in the flow through fraction. Taking into account the volume loaded onto the gel and the total fraction volume it can be estimated that less than 10% of the protein was retained on the column.

Analysis of the catalytic site mutant A4 EA-FLIS6 aliquots by immunoblot with an anti-FLAG antibody revealed that this protein was expressed to a much higher level than WT A4-FLIS6. This could be due to several reasons, notably the difference in total protein content in the extracts loaded onto the columns. However this was not corrected initially to avoid protein loss. The optimisations described were performed on WT A4-FLIS6; therefore optimal conditions for the mutant protein were not determined. The recombinant Baculovirus engineered to express A4 EA-FLIS6 may have a higher multiplicity of infection, or be more virulent than the wild type. Finally, it has consistently been observed with ADAR protein purifications that more protein is obtained with a catalytic mutant than a wild type protein preparation, although the difference is not usually as striking as that seen here (personal observation). However, the high level of A4 EA-FLIS6 expression highlights the inadequate binding of the protein to the DEAE Sepharose column as protein is present in all fractions analysed. This indicates that despite the large binding capacity of the DEAE Sepharose fast flow resin the interaction is not strong enough

to retain large amounts of recombinant A4-FLIS6 protein. This level of protein loss is too great to use DEAE Sepharose as the first step in A4-FLIS6 protein purification.

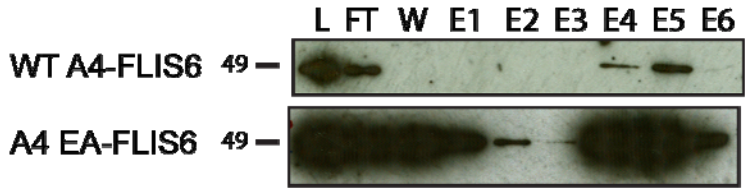


Figure 3.11: Purification of WT and catalytic site mutant APOBEC4 over DEAE sepharose

2x10^7 Sf9 cells were innoculated with recombinant Baculovirus directing expression of WT A4-FLIS6 (top panel) or a catalic site mutant A4 EA-FLIS6 (lower panel). Protein was purified by application of soluble extract to a DEAE Sepharose fast flow resin (Pharmcia Biotech). Resin was washed and protein was eluted with increasing concentrations of potassium chloride. Aliquots were taken and analysed by immunoblot using an anti-FLAG antibody. Unfortunately, only a fraction of the WT A4-FLIS6 bound to the DEAE Sepharose as a band was detected in the flow through. As observed previously the bound WT A4-FLIS6 was eluted with 500mM potassium chloride in elution fractions E4 and E5. By comparing band intesities it is estimated that less than 10% WT A4-FLIS6 was retained on the DEAE Sepharose matrix. Analysis of A4 EA-FLIS6 revealed this protein was expressed to a much higher level than the WT A4-FLIS6, despite being expressed in parallel cultures. The high expression highlights the lack of binding to the DEAE Sepharose as a band corresponding to A4 EA-FLIS6 can be detected in all fractions analysed. L = 1ul load (1ml total fraction volume), FT = 1ul flow through (1ml), W = 5ul wash (3ml), E1-E6 = 10ul elution fractions 1-6 (1ml).

As the most promising results were obtained with DEAE Sepharose, another stronger anionic matrix Macro-prep High Q Support (Bio Rad) was tested. Unfortunately very low amounts (approximately 5%) of recombinant A4-FLIS6 bound to this matrix, and protein was lost during the washes before a salt gradient was applied to elute protein (data not shown).

In addition to these purification procedures described I also followed a protocol published for the purification of recombinant human A3G from Baculovirus-infected Sf9 cells for small angle X-ray scattering experiments (Wedekind *et al.*, 2006). Briefly, cells were harvested 48 hours after infection and cell pellets were frozen at -70^{o}C. Cells were lysed by three freeze thaw cycles in liquid nitrogen and luke-warm water, followed by shearing through a 22 gauge needle. Detergent was added and nucleic acid degraded by addition of DNase and RNase. The sample was then incubated in 1M urea and centrifuged. The supernatant was incubated with Ni^{2+}

NTA agarose (Qiagen), washed with increasing concentrations of urea, and eluted with imidazole. Unfortunately this purification procedure failed to produce any soluble protein (data not shown).

Purified AID was only found to be active on ssDNA following treatment with RNase indicating the protein is capable of binding RNA tightly but cannot deaminate it (Bransteitter et al., 2003). In case A4-FLIS6 was binding nucleic acids which could alter its structure or interfere with binding to the resins, DNase and RNase was added to extracts and incubated at either 4^{o}C or 16 ^{o}C prior to purification. Unfortunately this did not increase binding to the matrices. As APOBEC/AID proteins are difficult to purify there is not a vast range of literature on the subject, however a paper published by Harold Smith detailed extensive recommendations for purifying APOBEC/AID proteins (Smith, 2007). Therefore factors thought to affect the purification of other APOBEC/AID proteins were tested. These included epitope tag purifications following attempts to break potential dimers or multimers with high salt conditions (2M potassium chloride or 1M sodium chloride), and purification at room temperature in the presence of Triton X-100 detergent. But these approaches were not successful as the majority (if not all) A4-FLIS6 protein was detected in the flow through fraction (data not shown).

In conclusion, attempts to purify a soluble C-terminal fragment of recombinant A4-FLIS6 from an *E.coli* expression system were not successful. Despite eventually obtaining soluble recombinant tagged A4-FLIS6 from Sf9 insect cells, purification proved to be extremely difficult. Taken together these results indicate that the soluble recombinant A4-FLIS6 expressed in Sf9 cells comprises a mixed population of protein in different conformations with different binding specificities. The epitope tags added to aid purification do not appear to be uniformly exposed, such that a maximum of only 20% of the soluble recombinant A4-FLIS6 could be purified. With no understanding of why there are differences in the protein produced there is no guarantee that the protein fraction purified would be representative of the biologically functional protein. Furthermore, the volume of cells required to produce enough soluble recombinant A4-FLIS6 with such inefficient purification would be

impractical. Therefore it was decided that this aspect of the project should not be continued.

Generation and purification of polyclonal antibodies to recombinant APOBEC4

To generate soluble A4 Alexey Ruzov subcloned recombinant A4-FLIS6 into the pGEX-6p-3 vector (GE Lifesciences) to produce a Glutathione S-transferase (GST) fusion protein in *E.coli*. The amino terminal GST-tag has a molecular weight of approximately 26kDa, which creates a fusion protein of approximately 71kDa hereafter called GST-A4-FLIS6. The presence of a large GST-tag made the recombinant protein soluble, and allowed purification using Glutathione sepharose (GE Lifesciences), however it was not used initially as the presence of such a large epitope tag may interfere with protein function. However, previous studies have successfully used GST-tagged APOBEC proteins purified from *E.coli* (Chester *et al.*, 2004). Recombinant GST-A4-FLIS6 was expressed in *E.coli*, purified using Glutathione sepharose and used to inoculate three guinea pigs for antisera production (Eurogentec). Sera from the final bleed was tested for its ability to detect recombinant A4-FLIS6 (Figure 3.12), all sera were able to detect the recombinant A4-FLIS6 antigen they were raised to.

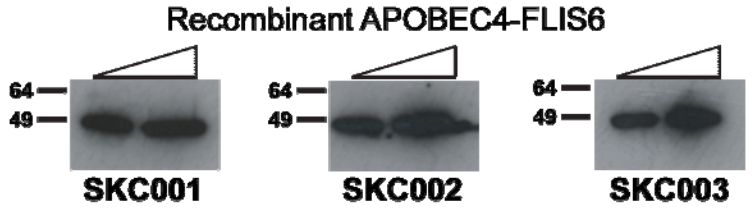


Figure 3.12: Testing APOBEC4 antibodies.
Recombinant GST-A4-FLIS6 protein purified from *E.coli* was injected into three guinea pigs to generate polyclonal antibodies. Unpurified sera here called SKC001, SKC002, and SKC003 were tested by western blot on two different concentrations (5µl and 10µl) of partially purified soluble recombinant A4-FLIS6 expressed in Sf9 insect cells. All primary antibodies were used at a dilution of 1:500, and secondary anti-guinea pig HRP was used at 1:1000 on all three blots. A band of approximately 45kDa was seen in all lanes indicating the antibodies were all capable of detecting recombinant A4-FLIS6 protein.

For affinity purification of the antisera raised to GST-A4-FLIS6 two purification columns were generated. Recombinant GST-A4-FLIS6 was expressed in large volumes (using conditions determined by A.Ruzov) and extracts of *E.coli* were made using the Bugbuster kit and protocol (Novagen). For purification Glutathione sepharose high performance (Amersham Biosciences) agarose beads were used, and recombinant protein was eluted with 20-50mM Glutathione. To remove antibodies that recognised the three epitope tags present on the antigen used for immunisation the A4 coding sequence was excised from the construct using the *SpeI* restriction enzyme and the vector was re-ligated to produce a recombinant GST-FLAG-6xHIS coding sequence, hereafter referred to as GST-FLIS6. This was also expressed and purified from *E.coli*.

Recombinant GST-A4-FLIS6 and GST-FLIS6 proteins were dialysed into coupling buffer to remove contaminating Glutathione, concentrated and then covalently bound to Cyanogen-Bromide-activated Sepharose (GE Healthcare). Of the three antisera two were purified over the GST-FLIS6 column first and the unbound fraction (flow through) was then loaded onto the GST-A4-FLIS6 column. The remaining antisera was purified in the opposite order, however this was not as successful. Initial experiments with affinity purified anti-A4 antibodies revealed all three could detect recombinant A4 protein expressed in either *E.coli* or Sf9 insect cells, however no bands were detected in mouse testis protein extracts. A4-null mice generated by Ian Adams were used as a negative control in these experiments. To determine the sensitivity of the antibodies the amount of recombinant GST-A4-FLIS6 protein present in the extract was quantified. Aliquots of known volume were analysed by SDS-PAGE with known amounts of bovine serum albumin (BSA). SDS-PAGE gels were stained with Sypro Ruby (Sigma) and quantified with ImageQuant software (GE Healthcare). A standard curve was generated from the BSA protein and used to calculate the concentration of full length recombinant GST-A4-FLIS6 (Figure 3.13). Importantly a much stronger band of lower molecular weight is seen on the Sypro Ruby-stained gel indicating degradation has occurred, but the full length protein was used for calculating concentration. From this gel it was estimated that 6ul extract contained 1ng full length recombinant GST-A4-FLIS6 protein.

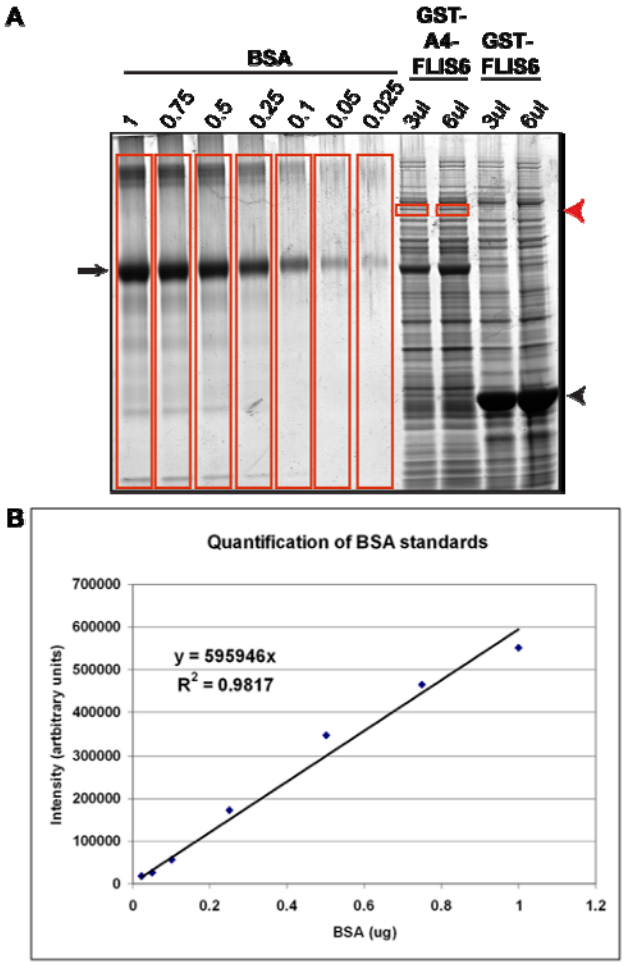


Figure 3.13: Quantification of GST-APOBEC4-FLIS6 protein.
A) Sypro-ruby stained SDS-PAGE gel showing serial dilutions of bovine serum albumin (BSA) standards used to estimate the concentration of recombinant GST-A4-FLIS6 in a known volume. Red boxes indicate areas used in intensity analysis, black arrow (left) indicates BSA protein, black arrowhead (right) indicates GST-FLIS6 control protein, and red arrowhead (right) indicates target GST-A4-FLIS6 protein. **B)** BSA intensities were used to generate a standard curve by plotting BSA concentration (ug) against Intensity (arbitrary units) and linear regression was used to determine the coefficient of the slope (here 595946). This graph was used to estimate the concentration of GST-A4-FLIS6 protein as 0.17ng/ul.

Numerous attempts were made to detect endogenous A4 in mouse testis extracts; however no bands could be detected. Finally, testis extracts from several different strains of mice were analysed by coomassie gel and immunoblot using anti-A4 SKC003 at a 1:100 dilution with known concentrations of recombinant GST-A4-FLIS6 (Figure 3.14). A large amount (100ug) of total testis extract was loaded to maximise the chance of detecting a small amount of protein. No endogenous A4

protein could be detected in 100ug total testis extract, where the affinity-purified anti-APOBEC antibody SKC003 could detect 1ng recombinant GST-A4-FLIS6, indicating if A4 protein is present in mouse testis it is at a lower amount than 1ng/100ug. Similar results were obtained with the two other affinity-purified sera. The epitope tag protein GST-FLIS6 was included as a control. The coomassie gel shows it was expressed at a significantly higher concentration than the A4 protein as there is a strong band present at approximately 30kDa (indicated by lower arrowhead). However purification of the antibody was completely successful as there is no corresponding signal present on the immunoblot. There is a higher molecular weight band present in the GST-FLIS6 sample that is recognised by the A4 antibody indicating contamination of this sample preparation has occurred with GST-A4-FLIS6 protein, this probably occurred during purification.

Another explanation for the lack of detection of endogenous A4 is that the protein used to immunize the rabbits was not in the same conformation as the endogenous protein. Given the problems detailed earlier with purifying recombinant A4 protein this may explain the lack of detection here. However, it would be highly unusual to have no antigen recognised with polyclonal antibodies.

Protein extracts were made from wild type mouse sperm and analysed by western blot but no bands were observed (data not shown). However, one explanation for the failure to detect endogenous A4 protein is that it is not expressed until after fertilization. Sperm mRNAs have been found in the embryo 3 hours post-fertilisation indicating that some mRNAs are delivered at fertilisation and not immediately degraded (Ostermeier *et al.*, 2004). However, due to problems collecting sufficient material, post-fertilization embryos could not be screened. A more likely explanation is that endogenous A4 protein is present at a level that is below the detection limits of the antibody. Given the problems with formation of insoluble inclusions when overexpressing the recombinant A4-FLIS6 protein it is likely that endogenous protein levels are maintained at a low level.

In addition to the three polyclonal antibodies generated in guinea pig, four custom peptide antibodies were raised in mouse (Eurogentec). All of the peptide antibodies detected recombinant A4 when analysed by western blot however the signal was considerably lower than the polyclonal antibodies. None of these antibodies detected endogenous A4 protein in testis samples (data not shown).

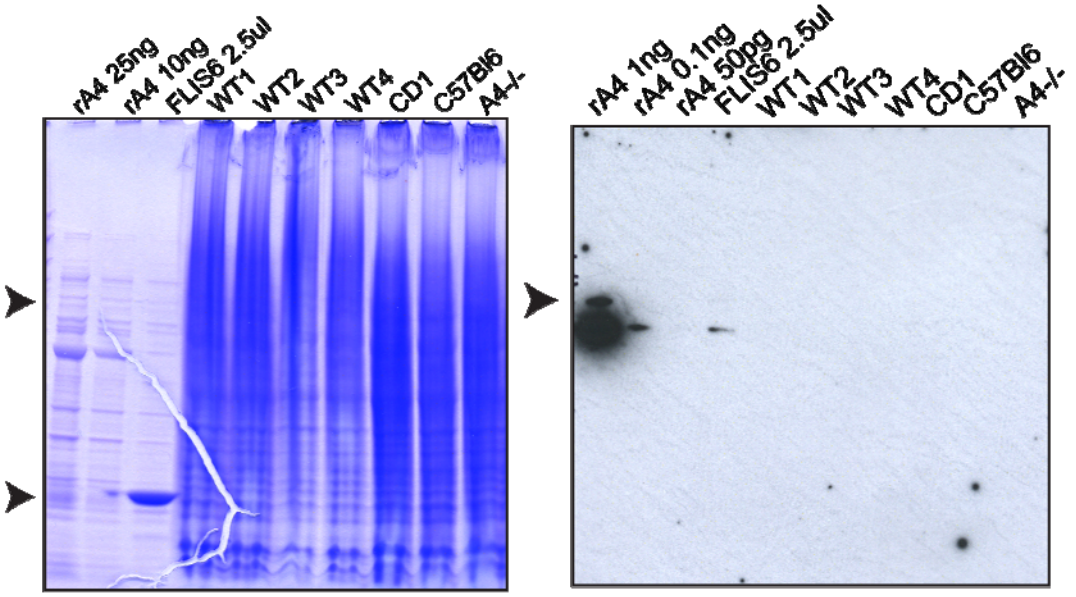


Figure 3.14: SDS-PAGE and Western blot analysis of APOBEC4 in testis. Testis extracts from several mice of different strains were analysed by coomassie stained SDS-PAGE (left) and western blot (right) for expression of A4. Testis extract from an A4-null mouse was included as a control. 100ug total testis extract was loaded with known concentrations of recombinant GST-A4-FLIS6 extract. Proteins appear smeared on the coomassie gel as high amounts were loaded to try to detect A4. No bands were observed on the western blot in the testis extracts of any of the strains analysed. Two bands of recombinant GST-A4-FLIS6 can be seen corresponding to 1ng full-length GST-A4-FLIS6 and a degradation product. When lower amounts of full-length protein were loaded (0.1ng and 50pg) only the degradation product was detected. A faint contaminating band corresponding to degraded GST-A4-FLIS6 is seen in the GST-FLIS6 lane indicating this protein extract was contaminated, however no band is seen at a lower molecular weight indicating the affinity purified antibody does not recognises A4 and does not recognise the GST-FLIS6 epitope tags.

Subcellular localisation of GFP-tagged APOBEC4

As there are no suitable cell lines derived from germline cells a heterologous expression system was used to determine the subcellular localisation of A4. Alexey Ruzov had previously shown that A4 expression was toxic in wild type Mouse

Embryonic Fibroblasts (MEFs; data not shown). The reason for this was not determined; however both the wild type A4 and the catalytic site mutant A4 elicited the same reaction. Therefore a p53-null cell line was used where the apoptotic cell death pathway initiated by p53 is no longer functional (Lande-Diner *et al.*, 2007). Wild type A4 and a catalytic site mutant A4 EA were cloned into a pEGFP expression vector under the control of a CMV promoter. Transient transfections were performed with lipofectamine (Invitrogen); cells were fixed in paraformaldehyde and imaged.

GFP-tagged A4 was predominantly cytoplasmic, however differences were observed with some cells showing diffuse cytoplasmic and nuclear staining (approximately 65-75% cells) with others displaying punctate cytoplasmic inclusions (approximately 25-35% cells; see Figure 3.16). These differences were attributed to different levels of protein expression whereby low levels of expression show diffuse cytoplasmic staining but higher levels of expression leads to aggregation of the protein into insoluble cytoplasmic inclusions as seen with other expression systems. The catalytic site mutant A4 EA displayed the same proportion of cells with diffuse cytoplasmic staining or punctate cytoplasmic inclusions. APOBEC/AID family members show a mix of nuclear and cytoplasmic localisation, which does not appear to correlate with antiretroviral activity, therefore we cannot infer targets of A4 from its expression pattern in a heterologous system. The localisation of AID/APOBEC family members correlates with size, such that smaller proteins show both nuclear and cytoplasmic localisation as seen here, whereas larger AID/APOBEC proteins display cytoplasmic localisation. This is attributed to the ability of smaller proteins to diffuse through the nuclear membrane, as may be occurring with A4. It is likely that these foci are insoluble aggregates of A4 but it cannot be ruled out that they could represent RNA-processing foci such as P bodies or stress granules which A3G has been shown to co-localise with (Wichroski et al., 2006).

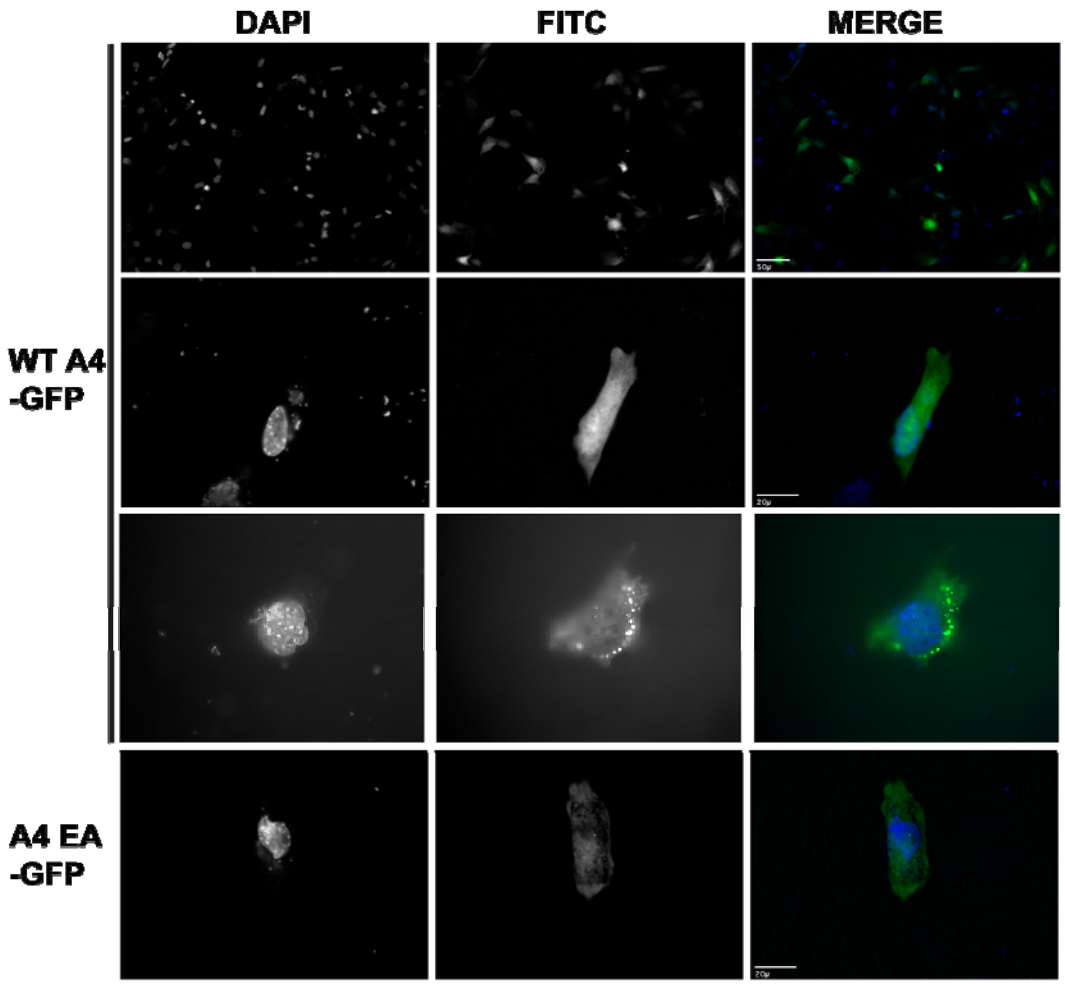


Figure 3.15: Localisation of GFP-tagged APOBEC4 in p53-/- MEFs.
p53-/- Mouse Embryonic Fibroblasts (MEFs) were transiently transfected with constructs expressing GFP-tagged wild type or mutant A4 under the CMV promoter. GFP is visualised using the FITC channel and DAPI-stained nuclei are visible in the DAPI channel. Both wild type and mutant A4 showed cytoplasmic staining, however different distribtion of signal was observed. Some cells exhibited diffuse nuclear and cytoplasmic staining (65-75%) and others showed localisation to discrete cytoplasmic inclusions (25-35%). The differences in localisation were attributed to differing amounts of protein expression, where at low levels diffuse localisation throughout the cell was observed but at high levels the protein forms insoluble aggregates as found in other expression systems.

p53-null MEFs transiently transfected with GFP-tagged A4 were used to analyse the specificity of the affinity-purified A4 antibodies. Immunofluorescence with a Texas Red conjugated anti-guinea pig IgG secondary antibody showed that anti-A4 antibody SKC003 signal appeared to co-localise with some of the punctate

cytoplasmic foci, however there was a high level background staining in cells that were not transfected, or transfected with a GFP construct alone (Figure 3.16). The non-specific background staining indicates these antibodies are not suitable for use in immunocytochemical analysis.

However, to confirm this result and determine whether these antibodies could detect endogenous A4 immunohistochemistry was performed on paraffin-embedded testis sections from wild type and A4-null mice. No specific staining was observed in the wild type testis and background staining was detected in both genotypes (data not shown). All three polyclonal antibodies were used for immunohistochemical analysis of paraffin-embedded testis sections however no signal could be detected that was absent in the A4-null mice indicating there was no specific staining. In addition to the three polyclonal antibodies generated in guinea pigs, four custom peptide antibodies were tested for their ability to detect both GFP-tagged A4 and endogenous A4 in mouse testis sections. However none of the antibodies analysed showed complete co-localisation with the GFP-tagged protein, and they all failed to detect a specific signal in the immunohistochemical analysis (data not shown). Different fixation methods (paraformaldehyde, alcohol based and Bouin's fixative) were tested with the guidance of Ian Adams, and two different detection methods were used (fluorescence-conjugated secondary antibodies and DAB detection) and antigen retrieval was also used (data not shown). However despite many attempts no specific signal A4 could be detected.

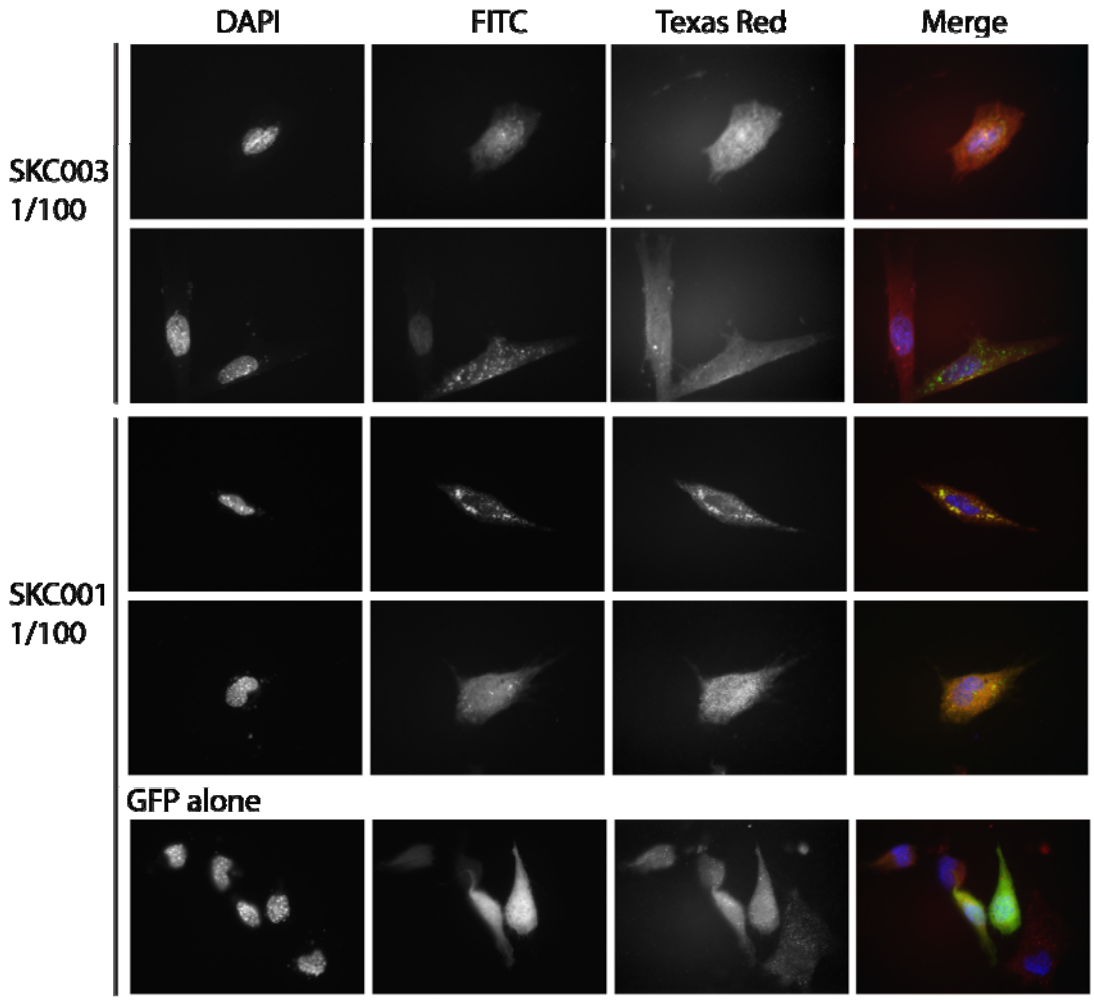


Figure 3.16: Immunofluorescence with anti-APOBEC4 antibodies .
p53-/- Mouse Embryonic Fibroblasts (MEFs) were transiently-transfected with a construct expressing either GFP-tagged A4 (top panels) or GFP alone (bottom panel) under the control of a CMV promoter. Cells were incubated with two different primary antibodies raised to A4 (SKC001 or SKC003) at a dilution of 1/100 and signal was detected using alexa 594-conjugated anti-guinea pig secondary antibody at 1/1000. GFP signal is visualised through the FITC channel, alexa fluor 594 is detected through the Texas Red channel, and DAPI-stained nuclei are visible in the DAPI channel. SKC003 (top two panels) did not show specifc staining as the punctate cytoplasmic GFP-positive oci were not detected in the Texas Red channel. SKC001 (lower three panels) showed the best co-localisation between the FITC and Texas Red channels as shown by detection of the cytoplasmic GFP-positive foci. However background staining was present in cells transfected with GFP alone, where there should be no signal. All images were taken using a 100x objective.

Taken together these results indicate that the affinity-purified antibodies can detect recombinant A4 overexpressed in several different systems when analysed by immunoblot. However, these antibodies are not suitable for immunocytochemistry. No endogenous A4 could be detected in mouse testis by immunoblot or immunhistochemistry; however I cannot conclude that it is not present as it may be expressed at a level that is below the detection limit of these antibodies. Further attempts were made to enrich endogenous A4 present in testis extracts by immunoprecipitation however no specific protein could be detected when comparing wild type and A4-null testis extract (data not shown).

E.coli DNA mutator assay of APOBEC4

To determine whether A4 could deaminate DNA an *E.coli* DNA mutator assay was performed as previously published (Coker *et al*., 2006; Harris *et al*., 2002) (Figure 3.17). Briefly, *E.coli* was transformed with a plasmid encoding either wild type A4 or catalytic site mutant A4 EA. The positive controls AID and A3G were also assayed along with the empty pTrc99a vector to give a background mutation frequency. Deamination events were determined by mutation of the *rpoB* gene which caused resistance to rifampicin, and a viable count was performed. The positive control A3G deaminated DNA with a median rate of 75 mutations per 10^9 cells (Figure 3.17). However the other positive control AID only displayed 12 mutations per 10^9 cells. This is much lower than the published data for AID activity in this assay, however there is considerable variation in the published data. One study reported AID median values of 103 mutants per 10^9 viable cells compared to a vector alone control which had 13 mutants per 10^9 viable cells (Petersen-Mahrt *et al.*, 2002), whilst another reported 180 mutants per 10^9 viable cells with a vector alone control frequency of 33 mutants per 10^9 viable cells (Harris et al., 2002). This gives a range of 5.5-7.9-fold enhancement of mutation frequency in the presence of AID, which is slightly higher than the 4-fold enhancement seen here. However the background mutation frequency is also higher in the published data. Harris and colleagues also assayed A3G which showed a median mutation frequency of 180-400

mutants per 10^9 viable cells (Harris et al., 2002), which is two-fold higher than the mutation frequency presented here.

However, A4 did not have a significant mutation rate above the background level of 3 mutants per 10^9 cells, indicating that in this assay A4 is not able to deaminate *E.coli* DNA. In addition there was no significant difference between wild type A4 and catalytic site mutant A4 which we would expect to see as catalytic activity would be required to generate mutations.

In addition to this assay a modified version was performed to analyse the ability of A4 to deaminate methylated DNA (Morgan et al., 2004). The modified assay involves transformation of a plasmid encoding the *SssI* DNA methyltransferase along with the plasmid encoding an AID/APOBEC protein, and selecting for both with ampicillin and kanamycin resistance. A modified strain of *E.coli* (ER1821, NEB) is required that is termed methylation-restrictive defective due to inactivation of four genes involved in restricting foreign methylated DNA (Kelleher and Raleigh, 1991). Transformation of either wild type A4 or catalytic site mutant A4 EA did not produce any colonies in this assay (data not shown). However the positive control for this assay is AID as it has been shown to deaminate methylated DNA (Morgan et al., 2004) and this also failed to give any colonies (data not shown), indicating that there was a problem with the assay or the AID clone used. In the absence of an antibody or epitope tags it was not possible to confirm expression of AID from this vector. Therefore no conclusion can be drawn about whether A4 is able to deaminate methylated DNA in *E.coli*.

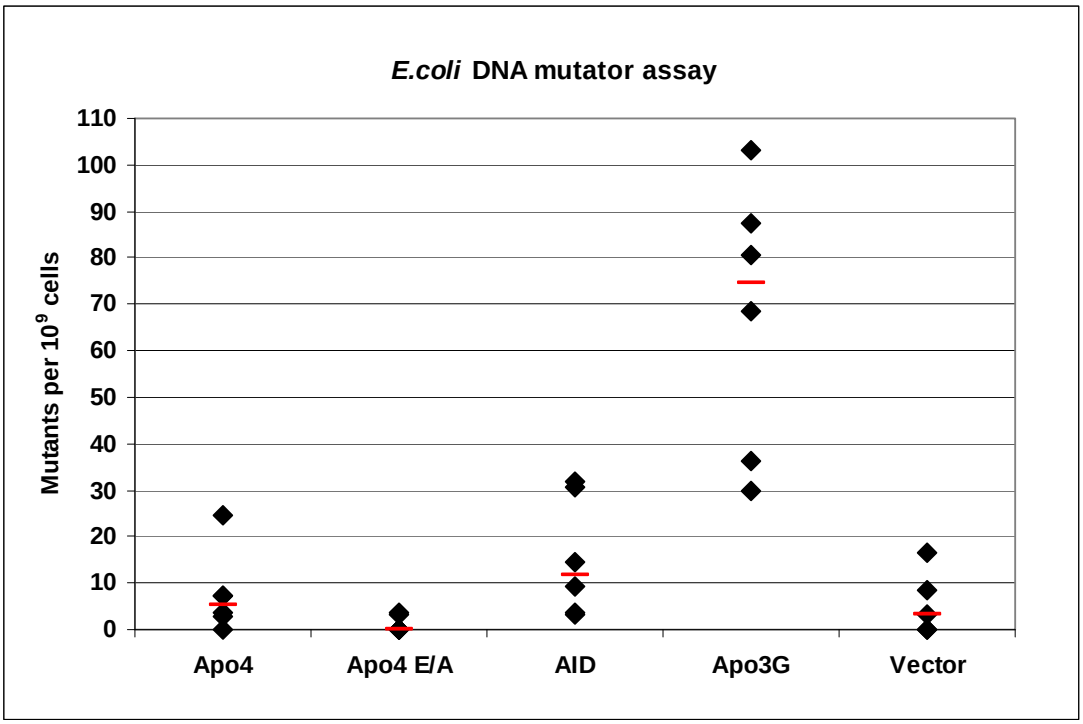


Figure 3.17: *E.coli* DNA mutator assay of APOBEC family members
E.coli KL16 cells were transformed with plasmids expressing wild type A4, catalytic site mutant A4 E/A, AID, A3G or the empty expression vector pTrc99a. Following induction with IPTG cultures were plated onto rifampicin plates or ampicillin plates to establish a viable count. Rif^R colonies were counted and plotted per 10^9 viable cells. A3G was capable of mutating *E.coli* DNA., whilst AID showed only a slight increase in mutation rate compared to vector control. A4 did not show a mutation rate above background. Black diamonds indicate independent replicates and red line indicates median value of 6 replicates.

3.3 Discussion

Detailed microarray profiling of both rodent and human spermatogenesis has revealed that conserved genes which are expressed early in spermatogenesis, either in somatic sertoli cells or mitotic spermatogonia are often expressed in other tissues and are therefore not specific for spermatogenesis (Chalmel *et al.*, 2007). In contrast, genes expressed at later stages of spermatogenesis, either during meiosis or post-meiotic spermatid stage are uniquely expressed in these highly specialised cell types. This microarray analysis confirmed that *Apobec4* is differentially expressed throughout spermatogenesis, with low expression detectable in spermatocytes and high expression in round spermatids, indicative of a specialised function in the post-meiotic spermatid (Chalmel *et al.*, 2007). However over 1,400 genes had a similar

expression profile including genes associated with the GO processes of fertilization and spermiogenesis. Following meiosis the sperm nucleus undergoes extensive remodelling when histones on the DNA are exchanged for protamines and this compact chromatin is transcriptionally dormant. Therefore transcription of genes encoding proteins required for spermiogenesis and fertilization occurs prior to chromatin compaction and mRNAs are stored in a translationally silent state until required. This means A4 protein expression may occur either at the round spermatid stage or later during spermiogenesis.

However, further work is required to determine whether *Apobec4* is regulated by miRNAs in general, and miR-741 or 742 in particular. Given that miRNAs can function in different ways it is of particular interest to determine whether A4 is regulated at the level of translation. This would provide an explanation as to why the *Apobec4* transcript is expressed during spermatogenesis but no endogenous A4 protein could be detected in mouse testis. As the expression pattern of miR-741 is similar to that of *Apobec4* mRNA it is an interesting hypothesis to suggest miR-741 acts to inhibit translation of A4 protein. Perhaps miRNA-741 regulates translation of *Apobec4* to maintain low levels of endogenous A4 protein. Expression of recombinant A4-FLIS6 led to aggregation of protein in insoluble inclusion bodies raising the possibility that the endogenous A4 levels may be kept low to prevent toxic aggregates from forming.

Whether mRNA is present in mature sperm is a somewhat controversial phenomenon with suggestions that any mRNAs detected arise from contamination, reviewed in (Miller *et al.*, 2005). However, without further investigation it cannot be excluded that *Apobec4* mRNA is present in mature sperm and delivered to the oocyte at fertilisation. The sperm mRNA may be degraded upon entry to the embryo; however there is evidence that some sperm mRNAs are present 3 hours after fertilisation indicating these mRNAs may have a function in the developing embryo (Ostermeier *et al.*, 2004).

Generating soluble recombinant A4-FLIS6 proved more difficult than expected, with initial *E.coli* and *Pichia pastoris* (M.Hogg, Masters Rotation Project) expression systems providing insufficient amounts for biochemical analysis. Amino-terminal truncation of A4 did not improve solubility as it did for A2. The Baculovirus Expression system generated more soluble recombinant A4-FLIS6, but the majority of the protein was still insoluble indicating A4 is an inherently insoluble protein. Other members of the AID/APOBEC family are also difficult to express and purify, and several groups have been trying to solve the crystal structure of A3 proteins for some time. However, much of the biochemical work on these proteins has been performed in cell culture models. This is not possible for A4 due to the lack of reliable germline cell culture models. Using a heterologous system for analysis of A4 is not ideal as co-factors may be required for function or regulation which may not be expressed in other cell types. Furthermore if the substrate of A4 is mRNA then these mRNAs may only be expressed during spermatogenesis.

A4 protein contains a putative coiled-coil domain at the carboxy terminus that is not found in other family members. This region may exacerbate problems with solubility and aggregation of overexpressed A4 protein as coiled-coil domains are often involved in protein-protein interactions. Removal of the coiled-coil domain may make purification easier, although this may alter biological function.

It has been hypothesised that the seclusion of recombinant APOBEC proteins in insoluble inclusion bodies serves a function to protect the *E.coli* cells from non-specific deamination that occurs when AID/APOBEC family members are expressed at high levels (Smith, 2007). If this is true then it could indicate that A4 may possess catalytic deaminase activity, however as it was also difficult to purify the catalytically inactive A4 protein this is unlikely to be the reason for the insolubility. In addition, the results from the *E.coli* DNA mutator assay do not support this hypothesis as A4 did not induce mutation above background levels. However, when A4-FLIS6 was overexpressed in *E.coli* for purification it formed insoluble aggregates and this may also have occurred when A4 was expressed in *E.coli* for the DNA mutator assay. If A4 was sequestered into insoluble aggregates it may not have

access to the *E.coli* DNA which would explain why no mutator activity was detected. However further work is required to determine whether a lower level of A4 expression would induce DNA mutations in this assay.

AID-induced mutation levels were significantly lower than published results in this assay indicating the controls for this assay may not have been optimal (Harris et al., 2002). It should be noted that published data from this assay varies greatly, however even the vector alone control used to assay background mutation frequency presented here is low by comparison to published results, indicating the assay conditions were not optimal. As mentioned earlier the absence of enhanced mutation frequency in *E.coli* expressing A4 may be due to aggregation of the A4 protein into insoluble inclusions. An alternative explanation for the low mutation frequency observed in the *E.coli* DNA mutator assay is that A4 does not deaminate DNA. This hypothesis is supported by the lack of a significant difference between the wild type and catalytic site mutant A4 in the *E.coli* DNA mutator assay. The A4 substrate could be RNA; this is supported by the fact that ADAR and ADAT proteins which deaminate RNA substrates contain multiple amino acids between the conserved cysteine residues (see alignment in Figure 1.2). In support of A4 deaminase activity of some kind it has been observed that when transformed, cultured and purified under the same conditions more protein is observed with a plasmid encoding the putative catalytic mutant of A4 (see Figure 3.11, and also personal observation from different expression systems). This has been observed in several different expression systems but until a function can be described for A4 this cannot be tested. In addition this phenomenon has also been observed for ADAR proteins, see Appendix I.

In his chapter on C-U editing that was published as this work was being undertaken, Harold Smith details the problems he encountered trying to purify and assay members of the AID/APOBEC family, specifically AID and APOBEC1 (Smith, 2007). This chapter was very useful as the extent of the solubility problems with other AID/APOBEC proteins was not apparent from published data. One of his first observations is that expression in *E.coli* is not recommended as these proteins are often sequestered into insoluble inclusion bodies. Furthermore he explains that once

purified, or even during purification, these proteins are liable to form insoluble aggregates which cannot be dissociated with 6M urea or 1% SDS treatment. This problem is apparently exacerbated by freezing, which is how all the protein I produced was stored during purification and analysis. The problems encountered in this project are not specific to A4, however much work on other family members has made use of cell culture systems to study the protein function and to avoid the processes of expression and purification. However a recent publication by Stenglein et al, (2010) characterises a novel function of A3 proteins in deamination and degradation of foreign DNA (Stenglein et al., 2010). This indicates cell culture systems which rely on transfection of plasmid DNA may not be ideal for studying members of the AID/APOBEC family as foreign DNA may be degraded leading to unstable expression of the plasmid-encoded protein.

The antibodies raised to recombinant A4 were able to detect 1ng recombinant protein by immunoblot. Affinity purification was successful as the antibody did not detect the epitope tags present in the antigen. However no signal was detected in 100ug testis extract indicating that if the protein is expressed in testis it is expressed at a low level. This cannot be ruled out as we would only expect to detect A4 protein in a subset of cells within the testis as indicated by the *Apobec4* transcript expression pattern. Further, given the problems with aggregation of overexpressed protein it is possible that the endogenous A4 levels are low. GFP-tagged A4 expressed in p53-null cells could be detected on an immunoblot by the affinity-purified A4 antibodies, indicating the antibody recognises the A4 protein when expressed in different systems with different epitope tags. However when analysed by immunocytochemistry there was little co-localisation of the GFP-tag and the antibody staining and considerable background staining was observed. Immunohistochemical analysis of paraffin-embedded testis sections from wild type and A4-null mice failed to demonstrate a specific A4 signal and attempts to enrich for endogenous A4 by immunoprecipitation were also unsuccessful.

The discovery that AID is able to inhibit LINE-1 retrotransposons through a deamination-independent manner is fascinating (MacDuff et al., 2009). Similar

inhibition of retroviral elements has been reported for A3 proteins that are localised in the cytoplasm (Bogerd et al., 2006a; Bogerd et al., 2006b; Esnault et al., 2005; Kinomoto et al., 2007; Schumacher et al., 2005; Stenglein and Harris, 2006). Interestingly, key catalytic residues in the active site were not required for retrotransposon inhibition (Stenglein and Harris, 2006). This enables one to envisage a model whereby A4 may have retained its ability to interact with retrotransposons in the germline, but may have lost catalytic deaminase activity due to the deleterious consequences of mutagenising the germline. In addition, the demonstration that A3G has a cellular role as an RNA-binding protein which can be relocated to stress granules and binds to transcribed retrotransposon RNAs (Chiu et al., 2006; Gallois-Montbrun et al., 2007; Kozak et al., 2006; Wichroski et al., 2006) offers another potential mechanism of A4 activity as an RNA-binding protein. However further work is required to determine whether A4 is capable of binding to RNA.

Unfortunately the technical problems with purification of A4 protein prevented biochemical characterisation of the protein. This, combined with the difficulty in detecting the endogenous protein, led to the decision that I should work on another project.

Chapter 4: Microarray analysis of *Adar*-null flies

4.1 Introduction

There is only one *Adar* gene in *Drosophila melanogaster*. The $Adar^{5G1}$ deletion encompasses the coding region and promoter of the *Adar* gene. *Adar*-null flies have decreased viability, approximately 20% of wild type levels, and they display decreased locomotor activity and develop age-dependent neurodegeneration (Palladino et al., 2000b). Expression of wild type *Adar* in the cholinergic neurons using the GAL4-UAS binary system is able to suppress the $Adar^{5G1}$ phenotype (Keegan et al., 2005). Intriguingly, expression of a catalytically inactive *Adar EA* transgene also under the *Cha-GAL4* promoter is able to suppress the age-dependent neurodegeneration but does not restore the locomotion defects or reduced viability (Leeanne McGurk, Thesis). Several transcripts which are known substrates of *Adar* have been implicated in locomotor activity, taken together with our earlier data this implies that there are site-specific editing events which are crucial for locomotion, which are absent in the flies expressing the catalytically inactive *Adar EA*. However, this also suggests that the age-dependent neurodegeneration observed in the $Adar^{5G1}$ flies is due to loss of a deaminase-independent function of the ADAR protein, as it is suppressed by expression of the catalytically inactive *Adar EA* transgene. This potential second function of ADAR is investigated further.

Ultrastructural investigation of the $Adar^{5G1}$ neurodegenerative phenotype using transmission electron microscopy performed by a previous PhD student in the lab revealed multiple membrane-bound vacuoles. These aberrant multilamellar whorls and vesicles are similar to those seen in lysosomal storage disorders including Tay-Sachs disease in humans. In *Drosophila* there are several mutants in components involved in carbohydrate storage that show similar phenotypes such as *benchwarmer* (Dermaut *et al.*, 2005) and *Swiss cheese* (Kretzschmar et al., 1997). The benchwarmer mutation has been mapped to a sugar transporter indicating this mutation models lysosomal storage disorders seen in humans. However, this phenotype could also be due to aberrant autophagy. Recently the *Drosophila blue cheese* mutant, which develops progressive neurodegeneration (Finley *et al.*, 2003), was shown to have impaired lysosomal trafficking (Lim and Kraut, 2009). A screen

for modifiers of the blue cheese phenotype identified components of the autophagy pathway (Simonsen *et al.*, 2007).

The catalytic function attributed to ADAR is RNA editing of A-to-I in dsRNA. However, recent work has shown that binding of ADAR protein to dsRNA can affect RNA processing, independent of deaminase activity. This was demonstrated in experiments investigating the effect of ADARs on micro-RNA processing, which showed that ADAR binding was capable of inhibiting cleavage by both DROSHA in the nucleus and DICER in the cytoplasm. This provides direct evidence that ADAR may have a role as an RNA-binding protein independent of its ability to deaminate adenosine residues (Heale et al., 2009). Further, as no catalytic activity has been described for mammalian ADAR3, elucidating a deaminase-independent function for *Drosophila* ADAR may shed some light on the function of ADAR3. This chapter aims to elucidate the deaminase-independent role of ADAR in the *Drosophila* brain.

To determine whether *Adar* had a more general role in RNA processing transcript levels were analysed by microarray. This revealed that a group of transcripts showed altered levels in the absence of *Adar*. Expression of either an active or inactive *Adar* transgene in the cholinergic neurons was used to isolate transcripts that showed "rescue" in response to *Adar* expression by searching for patterns of ADAR-regulated expression. This approach identified a subset of genes expressed in fly head that had altered expression indicative of regulation by the presence of ADAR. Cytochrome genes involved in response to oxidative stress were both upregulated and down regulated indicating ADAR is involved in the stress-response pathway. Transcript changes observed with the microarray analysis were confirmed using real time PCR analysis.

In vitro binding experiments were used to characterise the interaction of ADAR with novel transcripts identified as being regulated by ADAR according to the microarray data. This revealed ADAR is capable of binding many transcripts *in vitro* with a similar affinity to known substrates and therefore has the potential capacity to affect many transcripts through RNA-binding. To determine whether these effects could

also be seen *in vivo* an RNA-immunoprecipitation experiment was performed in *Drosophila* S2 cells overexpressing FLAG-tagged ADAR. *In vitro* editing assays were used to analyse whether the identified transcripts were substrates for deamination by ADAR.

ADARs have been shown to antagonise the RNAi pathway, leading to the hypothesis that antagonism of RNAi may explain the alteration in transcript levels seen in $Adar^{5G1}$ flies. To investigate this transcript levels were analysed in several different fly strains which have mutations in *Dicer-2*.

As *Adar* encodes an RNA-binding protein implicated in regulation of response to oxidative stress, an alternative approach to investigating the neurodegenerative phenotype observed in $Adar^{5G1}$ flies was employed through analysis of RNA oxidation levels. RNA oxidation has been shown to selectively occur in subsets of neurons which go on to develop pathological inclusions in several different neurodegenerative disorders. To determine whether levels of RNA oxidation differed in $Adar^{5G1}$ flies an *in vitro* immunoprecipitation was performed on RNA isolated from wild type and $Adar^{5G1}$ fly heads, with an antibody raised to 8-oxoguanosine. This method separates oxidised RNA from non-oxidised RNA for further investigation.

4.2 Results

Adar-null flies show locomotion defects and age-dependent neurodegeneration

As previously described *Drosophila melanogaster* lacking the *Adar* gene develop locomotion defects and age-dependent neurodegeneration (Ma et al., 2001; Palladino et al., 2000b). Further investigation of this phenotype demonstrated that it was due to lack of the *Adar* gene as the locomotion defects and age-dependent neurodegeneration could be suppressed by expression of an *Adar* transgene in the

cholinergic neurons, however expression of a catalytically inactive *Adar* EA transgene suppressed the neurodegeneration but could not alleviate the locomotion defects (Leeanne McGurk, Thesis). This implies that ADAR has an editing-independent function in the fly nervous system, and when this function is disrupted flies develop age-dependent neurodegeneration. To further characterise the neurodegenerative phenotype in $Adar^{5G1}$ flies and identify transcripts that are regulated by ADAR microarray analysis was performed. However, to ensure the changes observed were primary effects due to the lack of *Adar* rather than secondary effects induced by the neurodegenerative process flies were analysed at 5 days old.

Before further investigation of these phenotypes by microarray analysis, histological examination of 5 day old heads was performed on flies of four different genotypes: w^{1118}, $Adar^{5G1}$, $Adar^{5G1}$*:Cha-GAL4>UAS-dAdar*, and $Adar^{5G1}$*:Cha-GAL4>UAS-dAdar EA*. Haematoxylin and eosin staining of 8μm brain sections revealed that there was no evidence of neurodegeneration at five days, whilst holes in the brain are present in $Adar^{5G1}$ flies aged to 25 days (Figure 4.1). Large holes are observed in the mushroom body calyces which are involved in olfactory learning and memory processing (reviewed in (Heisenberg, 2003)). Small holes were occasionally observed in the rescue lines expressing active *Adar* or inactive *Adar EA* transgenes at the 25 day time point (arrowhead Figure 4.1) indicating the neurodegeneration is not completely suppressed but is significantly delayed by expression of ADAR in cholinergic neurons. Comparison of flies expressing wild type ADAR with flies expressing catalytically inactive ADAR EA in the cholinergic neurons revealed that flies expressing the catalytically inactive ADAR EA had fewer holes in the brain that wild type, which confirmed earlier observations made by a former PhD student Leeanne McGurk. This was attributed to differences in expression observed for the two transgenes, see Figure 4.4 and associated discussion.

No signs of neurodegeneration were observed at 5 days in the four lines analysed indicating any changes in transcript level observed should be due to absence of ADAR rather than secondary effects of neurodegeneration processes. A minimum of three brains were analysed per genotype.

Flies were genotyped at one day old and aged to five days without anaesthesia prior to analysis of locomotion. Locomotion defects were analysed by placing flies in a marked chamber for 2 mins and counting the number of times they crossed the lines. This assay was performed at 20^{o}C. Ten flies of each genotype were analysed and three measurements were taken per fly which were then averaged. Figure 4.1 shows the w^{1118} flies, here used as wild type, crossed the lines an average of 12 times in 2 mins, this was severely reduced in $Adar^{5G1}$ which showed 10-fold lower level of movement. This was significantly different from w^{1118} when analysed using an unpaired t-test ($p<0.0001$, $t=17.5$). The $Adar^{5G1}$ flies were obviously uncoordinated and often fell over onto their backs, spending large amounts of time trying to stand up. $Adar^{5G1}$ flies frequently died due to falling over and getting stuck in the food, although this could be prevented by frequently tipping flies onto fresh food. The locomotion defects were reversed by expression of a wild type *dAdar* transgene in the cholinergic neurons; however they were not rescued by expression of a catalytically inactive *dAdar* EA mutant transgene in the same system. Again, this difference was significant when analysed by unpaired t-test ($p<0.0001$, $t=10.92$).

This indicates that there are site-specific editing events lost in the $Adar^{5G1}$ flies which are essential for locomotor function. These sites are edited when active *dAdar* is expressed and not when the catalytic mutant is present. This is perhaps not surprising as many of the known substrates of *Drosophila melanogaster Adar* are voltage or ligand-gated ion channels. This demonstrates that correct editing of these targets is essential for correct locomotion in adult flies. These data also suggest that *Adar* has a second function independent of adenosine deamination, which when lost leads to neurodegeneration which is observed as holes in the brain. As *Adar* contains two RNA-binding domains that function independently of the deaminase domain, mechanisms of RNA processing were investigated further. Firstly I sought to determine whether ADAR regulated mRNAs levels in *Drosophila* brain. Microarray analysis was performed to identify transcripts that were altered in $Adar^{5G1}$ flies, and rescued by expression of an *Adar* transgene.

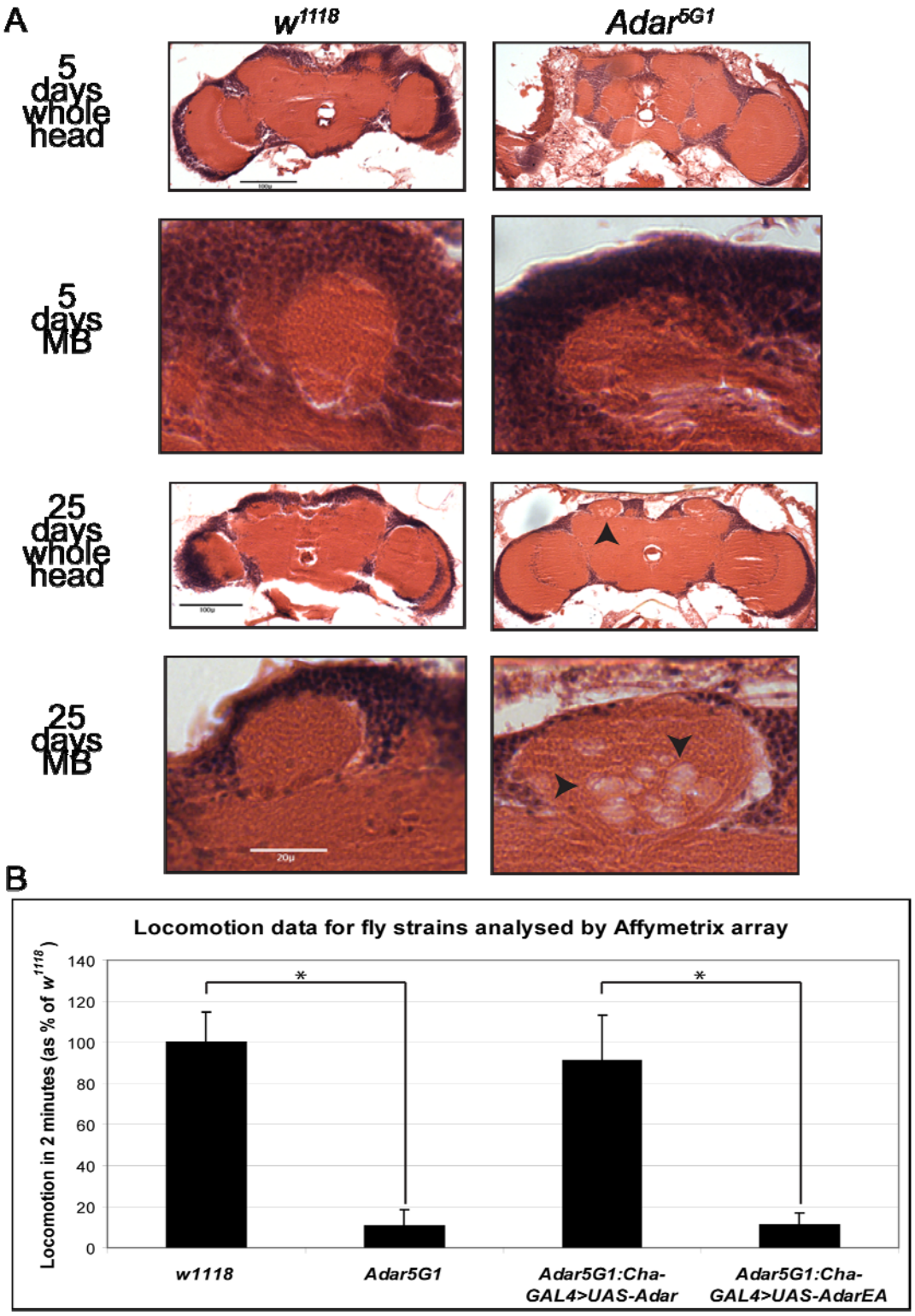


Figure 4.1: Characterising fly strains
Figure continued on following page with full legend

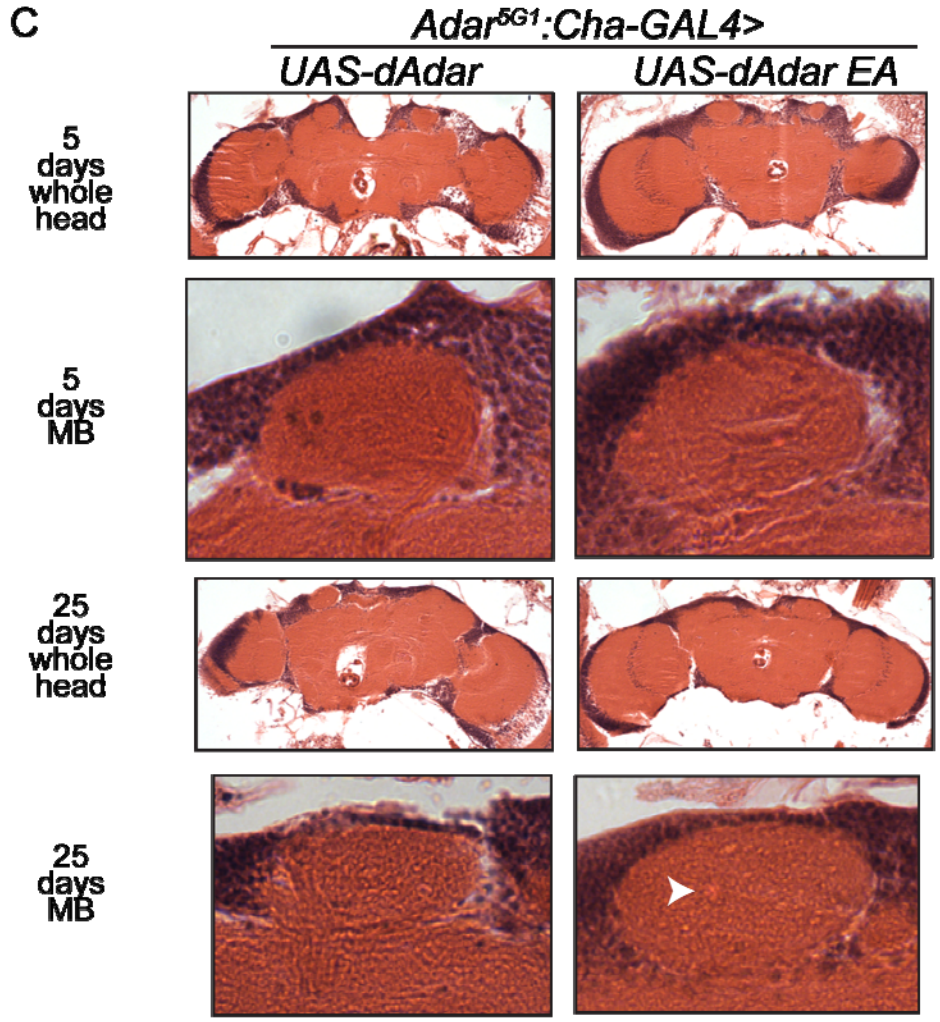


Figure 4.1: Characterising fly strains (continued).
Fly strains used in this study are w^{1118}, $Adar^{5G1}$, $Adar^{5G1}$*:Cha-GAL4>UAS-dAdar*, and $Adar^{5G1}$*:Cha-GAL4>UAS-dAdar EA*. The fly strains were analysed histologically at both 5 days and 25 days (panel A on the previous page shows w^{1118} and $Adar^{5G1}$, panel C above shows $Adar^{5G1}$*:Cha-GAL4>UAS-dAdar* and $Adar^{5G1}$*:Cha-GAL4>UAS-dAdar EA*) for evidence of neurodegeneration seen as holes in the Mushroom body (MB) calyces. None of the strains analysed showed evidence of neurodegeneration at 5 days. Large holes were apparent in $Adar^{5G1}$ flies at 25 days (black arrowheads), which were suppressed by re-expression of either an active *dAdar* transgene or a catalytically inactive *dAdar EA* transgene in the cholinergic neurons using the GAL4 binary system. However some small holes were observed in these rescue strains (white arrowheads), indicating expression of these transgenes in the cholinergic neurons is not sufficient to completely suppress neurodegeneration. Locomotion defects were also analysed in these fly strains (panel B, previous page). Flies were placed in a marked chamber and locomotion was analysed as the number of times they crossed the lines in two minutes was counted. Locomotion is shown as a percent of w^{1118} movement. $Adar^{5G1}$ flies showed a 10-fold decrease in locomotor activity when compared to w^{1118} flies, which was statistically significant with $p<0.0001$ (indicated by astersisk) when analysed by an unpaired t-test (t value = 17.05). This locomotion defect could be relieved by expression of a wild type *dAdar* transgene but not by a catalytically inactive *dAdar EA* transgene, which was statistically significant with $p<0.0001$ and a t value of 10.92.

Adar^{5G1} flies display reduced viability, estimated to be approximately 20% of *w*^{1118} levels. This viability defect is rescued by expression of an active *Adar* transgene in the cholinergic neurons; however it is not rescued by expression of the catalytically inactive *Adar EA* transgene. Therefore obtaining progeny from the *Adar*^{5G1} strain is difficult, and generating *Adar*^{5G1}*:Cha-GAL4>UAS-dAdar EA* progeny from crosses is even more difficult. For RNA extraction 50 fly heads were collected in TRIZOL (Invitrogen) and stored at -70^{o}C prior to extraction according to the manufacturer's protocol. In some experiments *Adar*^{5G1}*:Cha-GAL4>UAS-dAdar EA* RNA was not available due to problems obtaining flies of the correct genotype, therefore experiments were performed with *Adar*^{5G1}*:Cha-GAL4>UAS-dAdar* RNA.

Therefore there are four separate phenotypes associated with loss of ADAR expression, details are summarised in Table 4.1. The phenotypes are reduced viability, locomotion defects and uncoordinated movement, age-dependent neurodegeneration in the mushroom body calyces, and loss of site-specific editing of multiple adenosine residues in known substrates.

Genotype	Phenotype			
	Viability (%)	Locomotion (%)	Neurodegeneration	Editing
w ^{1118}	100	100	-	✓
Adar ^{5G1}	20	10.7	+++	✖
Adar ^{5G1} ***:Cha-GAL4>UAS-dAdar***	95	90.9	+	✓
Adar ^{5G1} ***:Cha-GAL4>UAS-dAdar EA***	20	11.3	+	✖

Table 4.1: Summary of phenotypes associated with deletion of *Adar*.
Summary of the four phenotypes associated with *Adar*^{5G1} deletion. Viability measurements are estimates based on flies collected for each genotype, locomotion was analysed as the number of times flies crossed the line in a marked chamber and data is given as a percent of *w*^{1118} levels, neurodegeneration is scored as present (+) or absent (-) based on histological examination, and editing is indicated as either present or absent based on sequence chromatograph analysis of known substrates (Jim Brindle, unpublished data).

Microarray analysis

To further characterise the defects in *Adar*^{5G1} flies microarray analysis was performed on RNA isolated from heads of five day old flies using the Affymetrix *Drosophila* Genome 2.0 array platform. This expression microarray contains 10-14

probes aligned to each 3'UTR to allow analysis of over 18,500 transcripts according to the Flybase database release 3.1. w^{1118} RNA was compared with $Adar^{5G1}$ RNA, and to detect transcripts that are regulated by ADAR RNA from flies expressing *Adar* in the cholinergic neurons was compared with RNA from flies expressing an inactive *Adar EA* transgene in the cholinergic neurons. Direct comparisons between w^{1118} flies and the rescue fly strains is not possible as in the rescue strains ADAR expression is restricted to the cholinergic neurons. In w^{1118} fly brain ADAR is expressed in the nucleus of all neurons (Leeanne McGurk, Thesis), whereas choline acetyltransferase expression is widespread but excluded from antennal primary sensory neurons (Yasuyama *et al.*, 1995). Importantly immunohistochemical analysis of choline acetyltransferase expression showed strong staining in the mushroom body calyces. However, direct comparison of expression patterns is difficult as ADAR is a nuclear protein whereas choline acetyltransferase is cytoplasmic.

Each sample was hybridised to the array individually and expression data was compared *in silico* to remove the need for dye swap replicates. The five day time point was chosen in an attempt to detect changes in transcript level due to lack of ADAR before neurodegeneration induced secondary effects. RNA was extracted with TRIZOL reagent according to the protocol and then further purified and concentrated over RNeasy MINelute kit (Qiagen) columns. Three replicates of each sample were hybridised to Affymetrix expression arrays at the Sir Henry Wellcome Functional Genomics Facility at the University of Glasgow. Initial analysis was performed as part of the service by Pawel Herzyk at The University of Glasgow to yield rank profile lists of changed transcripts when arrays were analysed in a pair-wise fashion (Breitling *et al.*, 2004). This analysis revealed that 364 genes were upregulated and 297 genes were down regulated with a false discovery rate (FDR) of less than 5%. As transcripts were both upregulated and down regulated this indicates that there is not a general affect occurring such as a lower level of transcription or increased turnover in $Adar^{5G1}$ flies. Of interest was the observation that some transcripts changed dramatically when comparing w^{1118} and $Adar^{5G1}$. In $Adar^{5G1}$ flies

the top-changing transcripts showed 18-fold down regulation (*CG11205, photolyase*) and 46-fold up regulation (*CG14204*) respectively.

However this analysis was not very useful for comparing one transcript across all fly strains. Therefore I performed a more comprehensive analysis using the freely available DNA-Chip Analyzer (dChip) software (Li and Wong, 2001). CEL files were converted into dChip files and MAS normalisation was performed using the array which displayed the median intensity value (10_dAdar_02; 129). I then searched the array data for patterns; specifically I wanted to isolate genes that were regulated by *Adar*. To do this I searched for genes that were mis-regulated in $Adar^{5G1}$ flies and "rescued" with re-expression of either the active or the inactive *Adar* transgene. If ADAR has a secondary function separate from RNA editing this would be present in both the active and inactive "rescue" lines; therefore I searched for the two patterns described in Table 4.2.

	w^{1118}	$Adar^{5G1}$	$Adar^{5G1}$:Cha-GAL4>	
			UAS-dAdar	UAS-dAdar EA
Upregulated	baseline	up	down	down
Downregulated	baseline	down	up	up

Table 4.2: Patterns used for analysis of microarray data.
Where w^{1118} was used as a baseline to compare other strains to. Colours correspond to the colours used in the heat maps in Figures 4.2 & 4.3 showing results of the searches, where upregulated is red and downregulated is green.

Searching for upregulated transcripts revealed a list of 59 genes with and a FDR of 0%, as estimated using dChip to search 100 permutations of these search criteria. A heat map showing these 59 genes clustered by expression profile is shown in Figure 4.2. A comprehensive list of the gene expression values and fold changes is shown in Appendix II, Table AII.1. Similarly a search for downregulated transcripts produced a list of 73 genes with a p value of less than 0.05 and a FDR of 0%, which are displayed as a heat map of expression data in Figure 4.3, and listed in detail in Table AII.2. One caveat for these searches is that transcript levels do not need to be restored to w^{1118} levels in the "rescue" lines; they only need to be significantly different from $Adar^{5G1}$ in the direction specified. These refined lists of genes should

comprise a list of transcripts that are affected by the presence of *Adar*, indicative of regulation by ADAR. Importantly transcripts were identified that were upregulated and downregulated when comparing w^{1118} with $Adar^{5G1}$ flies, indicating there is not a general lower level of transcription in *Adar*-null flies.

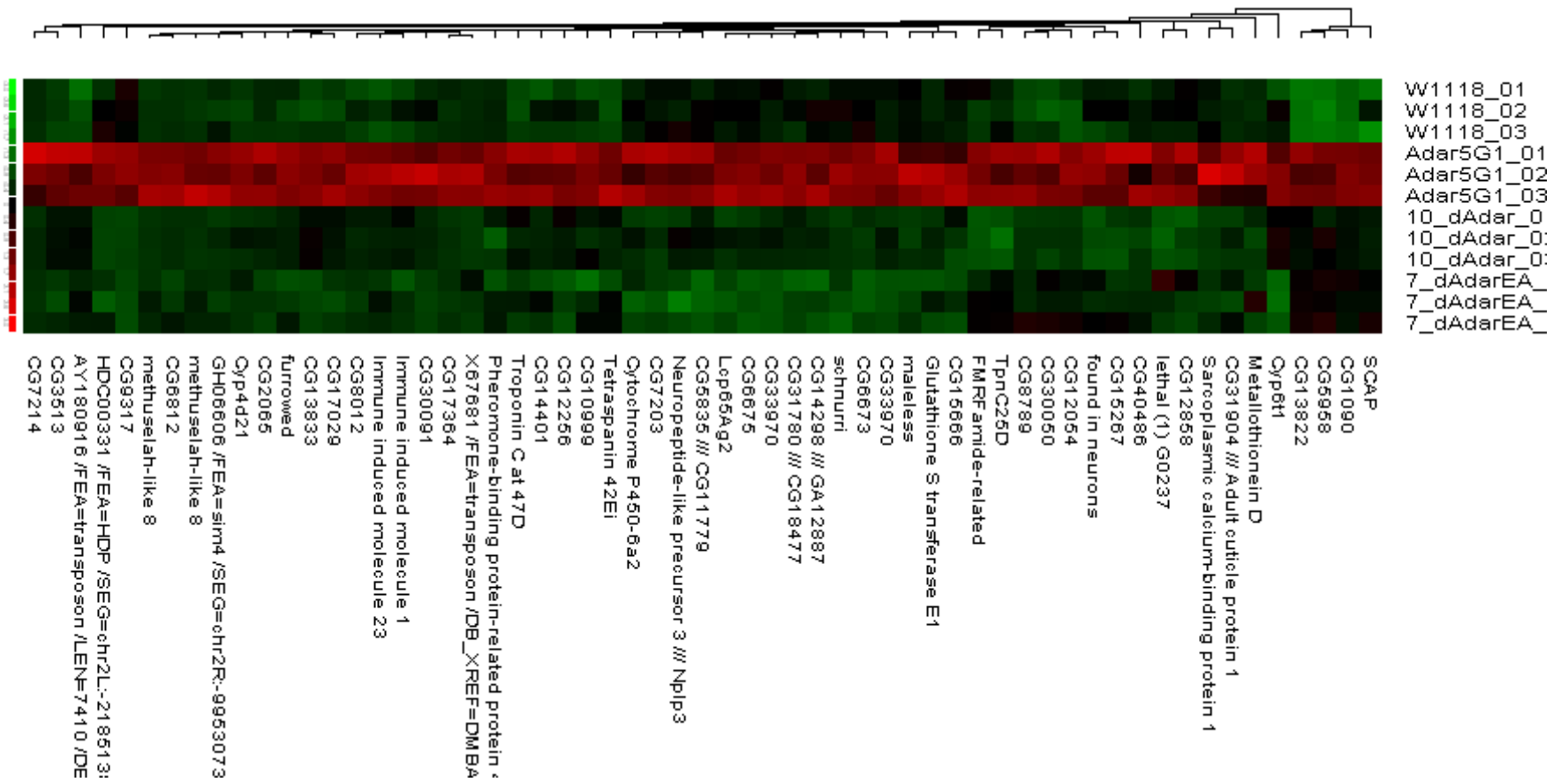


Figure 4.2: Heat map of results from a search for upregulated transcripts in the microarray data.
The 3 individual replicates for each genotype are listed on the right. Genes are clustered according to similar expression profiles.

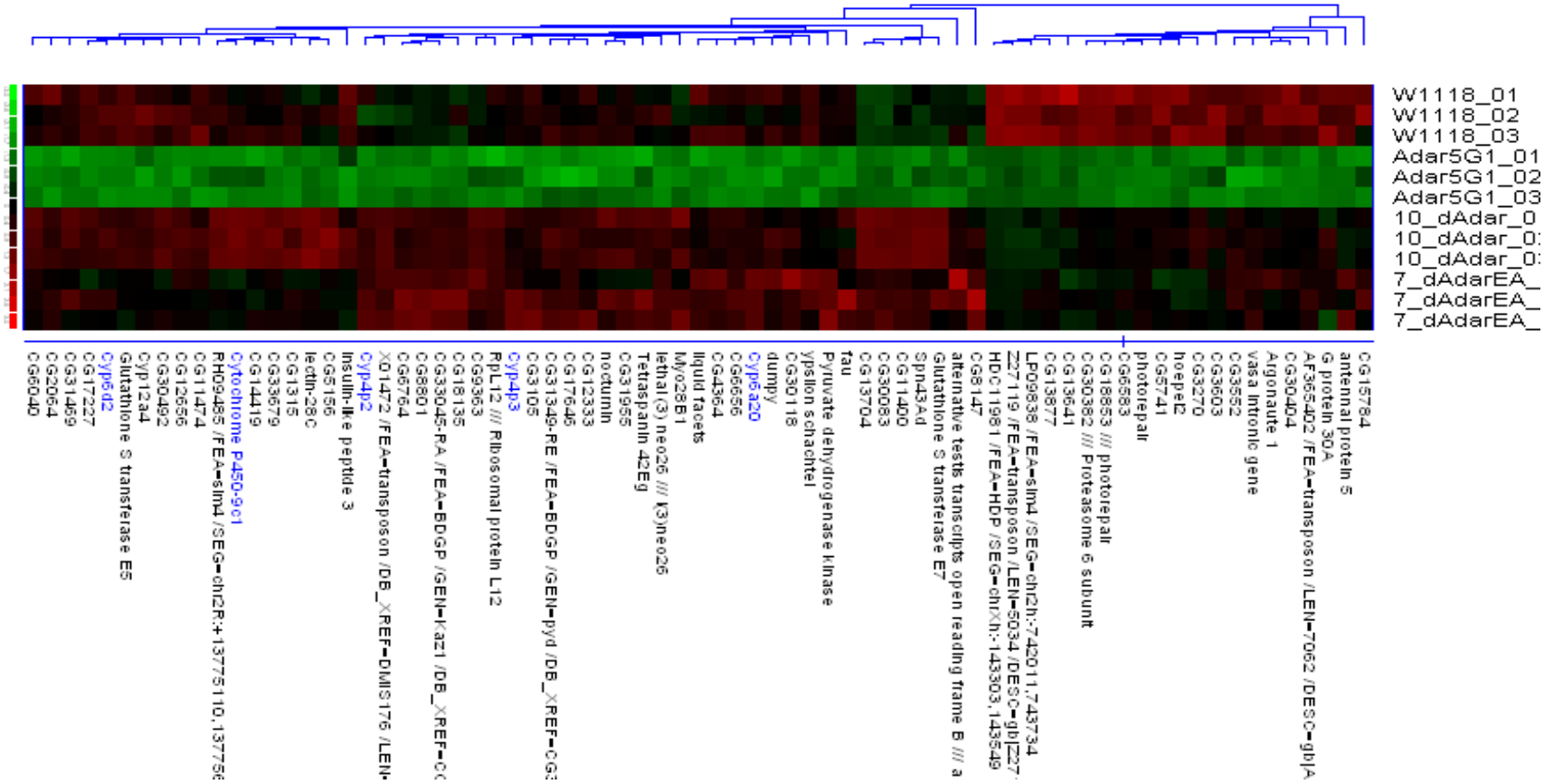


Figure 4.3: Heat map showing results from a search for donwregulated transcripts in the microarray data.
The 3 individual replicates for each genotype are listed on the right. Genes are clustered according to similar expression profiles.

The *Adar* transcript was absent from the downregulated list. However investigation of *Adar* transcript level by real time PCR revealed that expression of the active *Adar* transgene in the cholinergic neurons is very low, such that it would not be detected by the specified search parameters (Figure 4.4). However, given that this level of expression is sufficient to suppress the neurodegeneration and locomotion defects these parameters were not changed. Normally *Adar* expression is highly regulated and high levels of expression are detrimental to organisms, indeed expression of a highly active *Adar S* transgene under an *Actin-GAL4* promoter is lethal (Keegan et al., 2005). Therefore this low level of expression from the transgene is expected.

What is unexpected is the high level of expression of the catalytically inactive *dAdar EA* transgene. This could be due to several reasons. Firstly, the transgenes are inserted randomly in the genome and could be under the influence of neighbouring gene regulatory signals. Or secondly, hyper-editing seen with high levels of active *Adar* is detrimental to the organism and therefore selected against so expression of active *Adar* is usually low. However the catalytic mutant is not capable of editing RNAs; therefore this low level of expression may not be necessary with the catalytic mutant. The second hypothesis is supported by observations that the mutant ADAR protein is consistently expressed at a higher level that the wild type ADAR protein when overexpressed in yeast for biochemical analysis (personal observation, data not shown). Further support comes from the observation that three independent lines of each genotype were generated and analysed in a locomotion assay (Liam Keegan and Jim Brindle, unpublished data). Similar results were obtained for all three lines indicating regulatory signals from genes surrounding the insertion site are not contributing to the expression level.

Interestingly, expression of the inactive transgene in the cholinergic neurons produces a better suppression of the neurodegenerative phenotype than the wild type ADAR. This is likely to be partly due to the high level of expression of this transgene; however human ADAR1 is also regulated post-translationally via sumoylation (Desterro *et al.*, 2005).

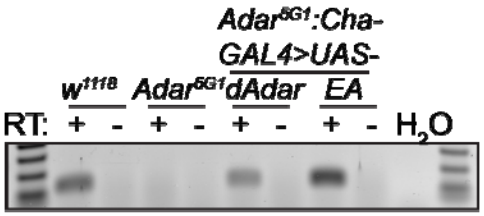


Figure 4.4: RT-PCR of *Adar* transcript

Adar **transcript levels were analysed in fly strains used for microarray analysis. Primers were designed for real time PCR and produce a ~150bp intron-spanning product. *Adar* transcript is present in w^{1118} flies and absent from $Adar^{5G1}$ flies. $Adar^{5G1}$ flies expressing wild type *dAdar* in the cholinergic neurons show lower levels of *Adar* than w^{1118}, however the catalytic mutant *Adar EA* transgene is expressed at higher levels than wild type. Real time PCR results are in Figure 4.5**

Unfortunately immunoblot analysis of ADAR protein levels in these transgenic strains using either an anti-human *Adar2* antibody which cross reacts with *Drosophila* ADAR or an anti-FLAG antibody was not successful, indicating the endogenous protein and the expression from the transgenes is low *in vivo*. To definitively quantify the protein levels an antibody specific to *Drosophila Adar* is required, this would allow direct comparison of expression patterns and levels between the transgenic strains and wild type levels in w^{1118} flies.

Analysis of the microarray data confirmed earlier work to define the $Adar^{5G1}$ mutation (Leeanne McGurk, Thesis). TAIL-PCR was performed to define the breakpoints of the $Adar^{5G1}$ mutation and revealed that the deletion removed 36.5kb of genomic sequence surrounding the *Adar* gene on the X chromosome. The deleted region contains *Adar* and two additional genes *CG14809* and *CG38206*, which are not known to be expressed in the brain. However, sequence analysis revealed that P element mobilization used to generate the deletion had also generated a small inversion with a breakpoint within a neighbouring gene called *Sarcoglycanδ* (*CG14808*). Neither of these genes are annotated as being expressed in the CNS, however analysing the microarray data it can now be confirmed that these transcripts are not altered in the RNA extracted from $Adar^{5G1}$ fly heads.

Two lists of genes that were altered in the microarray were generated, those that were upregulated in *Adar*^{5G1} flies (59 in total) and those that were down regulated in *Adar*^{5G1} flies (72 in total). These lists were analysed further using the functional enrichment of gene ontology tool FatiGO from BABELOMICS (Al-Shahrour *et al.*, 2006). Transcripts identified from both pattern searches were compared to the rest of the *Drosophila* genome (14,703 genes) to determine whether any gene ontology (GO) terms were significantly over-represented. The list of over-represented GO terms is shown in Table 4.3. Terms are classed according to cellular component or location, molecular function, or conserved domains as defined by Interpro database. The number listed with the GO term indicates the level of the term within a hierarchical cluster, where lower numbers indicate general processes and as the number increases the specificity increases. Therefore some processes appear multiple times, indicating both general terms associated with the process and more specific terms are significantly over-represented.

The GO analysis revealed that cytochrome p450 genes were over-represented in the *Adar* downregulated list, both in GO terms such as "mono-oxygenase activity" and cellular compartments like "microsome" and in Interpro domains. However, although not statistically significant cytochrome-associated genes were also present in the upregulated list indicating a general de-regulation. Studies into transcriptional responses to both paraquat stress and hydrogen peroxide stress have shown a similar altered regulation of cytochrome genes (Girardot *et al.*, 2004). The downregulated list showed GO terms associated with cytoplasmic membranous structures, which could represent components involved in synaptic transmission. The upregulated list showed fewer statistically significant terms, with cytoplasmic membrane-associated terms showing significance. Surprisingly, the upregulated list also showed over-representation of GO terms related to flagella. This was not expected; however a *Drosophila* mutant in the *sas-4* centrosomal protein has defects in flagella formation and demonstrates severe uncoordination due to lack of sensory neuron cilia involved in mechanosensation (Basto *et al.*, 2006). The *sas-4* mutants hatch but die shortly afterwards, often getting stuck in the food due to the uncoordinated movement. *Adar*^{5G1} flies frequently fall over and often get stuck in the food; however it is

possible to keep them alive by frequently changing the food indicating this phenotype is not as severe as the *sas-4* mutants.

However, as cytochrome genes have been shown to be regulated in response to oxidative stress it cannot be ruled out that this de-regulation may be a response to initial stages of neurodegeneration occurring rather than a causative change. Cytochrome genes have also been shown to be upregulated in response to the accumulation of a substrate, although this data was generated in humans with reference to drug metabolism this could be extrapolated to indicate the presence of an endogenous toxic product in the brains of *Adar*-null flies. Interestingly Glutathione S transferase genes were also found on both the lists although not significantly over-represented, these have also been shown to be regulated in response to oxidative stress (Girardot et al., 2004).

Class	GO term	Adjusted p-value
	Down regulated genes	
Cellular component	4: cell fraction (GO:0000267)	0.00174
	5: membrane fraction (GO:0005624)	0.00195
	6: vesicular fraction (GO:0042598)	0.00174
	7: microsome (GO:0005792)	0.00189
	7: cytoplasmic part (GO:0044444)	0.01906
Molecular function	3: tetrapyrrole binding (GO:0046906)	0.00210
	4: monooxygenase activity (GO:0004497)	0.00163
	4: heme binding (GO:0020037)	0.00163
Interpro domains	IPR001128: Cytochrome p450	0.00966
	IPR002401: Cytochrome p450, E-class, group I	0.00966
	IPR002403: Cytochrome p450, E-class, group IV	0.00966
	IPR002402: Cytochrome p450, E-class, group II	0.02202
	Up regulated genes	
Cellular component	4: cell projection (GO:0042995)	0.04635
	5: flagellum (GO:0019861)	0.01076
	5: membrane fraction (GO:0005624)	0.04913
	6: flagellin-based flagellum (GO:0009288	0.01169
	7: microsome (GO:0005792)	0.04816

Table 4.3: Analysis of enriched GO terms associated with transcripts identified as altered according to microarray analysis.
Gene Ontology (GO) terms and identifiers associated with the list of 72 down regulated genes and 59 upregulated genes. GO terms are grouped into cellular component, molecular function, and Interpro domains and within these classes GO terms are organised in hierarchical levels 1-9 where 1 indicates general terms and 9 indicates specific terms. No significantly over-represented GO terms associated with molecular function or Interpro domains were found in the upregulated gene list. Adjusted p-values are listed, all terms

listed have a p-value <0.05. Data compiled using the FATiGO software on the BABELOMICS website (Al-Shahrour *et al.*, 2006).

The *hypnos-2* mutant which is a deletion in *Adar,* was originally discovered in a screen for mutants that show increased susceptibility to oxygen deprivation (Ma et al., 2001). The authors hypothesised that this was due to the lack of site-specific editing of ion channels in the absence of *Adar*, which prevents adaptation to environmental stresses such as low oxygen levels. Therefore in the absence of *Adar* flies were more susceptible to oxygen deprivation and neuronal degeneration. Transcript analysis in *hypnos-2* mutants showed that several cytochrome genes were upregulated (Chen *et al.*, 2004), further these genes were downregulated in flies overexpressing ADAR indicating they are regulated by ADAR. However, the cytochrome genes identified by Chen et al (2004) are different genes from those identified here (Chen et al., 2004). This correlates with a general de-regulation of genes involved in response to oxidative stress. The microarray data presented here indicates that *Adar* may regulate transcripts involved in the response to oxidative damage, in a manner independent of site specific editing as levels of these transcripts are restored by expression of a catalytically inactive *Adar* transgene.

ADAR2 has also been implicated in neuronal cell death following forebrain ischemia (Peng et al., 2006). Elegant experiments in mouse have demonstrated that *Adar2* expression is down regulated following ischemic insult in vulnerable neurons, which leads to neuronal degeneration due to calcium influx. Interestingly, the transcription factor responsible for *Adar2* regulation in this system, CREB, was shown to be upregulated in resistant neuron populations indicating *Adar2* may be part of a protective response in these neurons. There is a parallel in *Drosophila* as in the absence of *Adar* flies are more sensitive to oxygen deprivation, similar to the vulnerable CA1 pyramidal neurons in mouse, and cytochrome genes are de-regulated. Intriguingly the *hypnos-2 Adar* mutant was shown to have increased resistance to paraquat treatment, a compound that is used to induce oxidative stress through the generation of reactive oxygen species (Chen et al., 2004).

Importantly, this kind of analysis depends entirely on how well the genome and the genes of interest are annotated. Of the list of 131 genes analysed 81 (61.8%) had a molecular function GO term attributed to them, and 94 (71.8%) contained a functional motif as recognised by the Interpro integrated database. This can be compared to 60.8% and 73.5% for the rest of the genome annotated with molecular function GO terms and Interpro domains respectively. These numbers indicate the annotation levels are comparable for the lists being analysed and the genome annotation, however it also indicates that 30-40% of the genes identified have no function attributed to them making comprehensive analysis difficult.

Confirmation of microarray results by real time PCR

Transcripts identified as being altered in the microarray data were analysed by real time PCR to confirm the changes (Figure 4.5). Primers were initially designed to the body of the gene however to accurately measure the same region of the transcript as the microarray probes, primers were also designed to the 3'UTR region. Initial analysis was performed with *Actin* and *GAPDH* reference genes, however as results were generally consistent further analysis was performed with *GAPDH* for normalisation. Graphs show the average fold change as calculated from a minimum of two biological replicates and two technical replicates with error bars indicating the standard deviation. Analysis of *Adar* transcript levels revealed a 10-fold decrease in $Adar^{5G1}$ flies; however some signal was still detected. This could be due to the technique of real time PCR which rarely gives an absolute zero value, although a low level of contamination with wild type RNA cannot be excluded. *Adar* expression is known to be low in w^{1118} flies, meaning the fold change is unlikely to be large.

Rab3 GEF was investigated as initial results from microarray analysis of w^{1118} and $Adar^{5G1}$ showed it was 10.9-fold downregulated in $Adar^{5G1}$, and it is known to be edited in the 3'UTR region by ADAR (Stapleton et al., 2006). *Rab3 GEF* was an interesting candidate for investigation as may be regulated in response to oxidative stress (Zou *et al.*, 2000). However real time PCR analysis of RNA extracted from flies expressing *dAdar* and *dAdar EA* transgenes showed that *Rab3 GEF* transcript

levels were not restored (Figure 4.5), which confirms the results obtained by microarray analysis of the rescue strains. As the expression of *Rab3 GEF* is not significantly different in the rescue lines from the expression in *Adar*^{5G1} flies, this indicates its activity is not linked to the phenotypes of locomotion defects or age-dependent neurodegeneration. One explanation for this is that *Rab3 GEF* may be expressed and regulated by ADAR outside the cholinergic neurons. *Rab3 GEF* is part of a conserved pathway involved in synaptic vesicle recycling at the neuromuscular junction and mice lacking the *Rab3 GEF* homolog *Rab3 GEP* die postnatally with 10-fold fewer synaptic vesicles (Tanaka *et al.*, 2001). However, localisation of *Drosophila* Rab3 GEF is required to determine whether it is expressed outside the cholinergic neurons.

Photolyase (*Phr*, also known as *photorepair*) transcript level was dramatically altered in *Adar*^{5G1} flies when analysed by both microarray (18-fold downregulated) and real time PCR (10-25 fold downregulated depending on primer pair used, see Figure 4.5). *Photolyase* is an enzyme that catalyzes the repair of UV induced pyrimidine dimers using energy from white light. Mutational analysis originally isolated a recessive allele that was defective in both photorepair and base-excision repair indicating a dual or common role between these pathways (Boyd and Harris, 1987). Interestingly, the 3' region of the *photolyase* coding sequence is present in two copies in the *D.melanogaster* genome. This duplication likely arose through either large-loop mismatch repair or replication slippage as the fused gene and the parental genes lie within 2 annotated genes of each other (Rogers *et al.*, 2009). One complete copy is present at locus *CG11205*, and a partial duplication has produced another gene with the same 3'UTR region (*CG18853*) which is itself a fusion product with another gene. As such, primers or probes designed to the 3' end of the transcript will detect RNA from both genomic loci. Therefore primers within the body of the gene were used to analyse the levels of photolyase (*CG11205*) by real time PCR. Analysis of *CG18853* transcript levels was not possible as the 5' region of this gene is shared with another gene, therefore specific primers could not be designed.

Drosophila Insulin-like peptide 3 (*Ilp3*) was also found to be down regulated in *Adar*^{5G1} flies and expression was recovered with expression of *dAdar* and *dAdar EA* transgenes, although not to the same level as transcripts in *w*^{1118}(Figure 4.5). *Ilp3* expression was 4.5-fold downregulated according to the microarray and 10-fold downregulated according to real time PCR analysis. *Ilp3* is one of seven insulin-like signalling molecules identified through searching the *Drosophila* genome for residues that are conserved in insulin. *Ilp3* is unique among these seven genes for showing the most restricted expression pattern, being expressed in only seven cells within each hemisphere of the developing larval brain, which are known as median neurosecretory cells (Brogiolo *et al.*, 2001). *Ilp3* is one of three ILP's expressed in adult brain, yet there is only one insulin receptor. Flies lacking 5 of the *Ilp* genes (*Ilp1-5*) that are clustered on chromosome III are small with poor growth and viability (Zhang *et al.*, 2009). However deletion of the one insulin receptor gene (*DInR*) is lethal (Chen *et al.*, 1996), indicating compensation by *Ilp6 & 7* is occurring in the *Ilp1-5*-null flies. Interestingly expression of *Ilp3* and *Ilp5* is regulated by nutrient availability such that these transcripts are not produced under starvation conditions (Ikeya *et al.*, 2002). *Adar*^{5G1} flies may have problems feeding due to lack of coordinated movement and long periods lying on their backs; however we might expect both *Ilp3* and *Ilp5* to be altered if starvation were the cause of altered gene expression. Characterisation of single *Ilp* mutants has revealed that there is functional redundancy and compensation of expression (Gronke *et al.*, 2010). Indeed we observe a slight upregulation on *Ilp7* in our microarray analysis (1.38-fold, 16% FDR), although this difference is too low to be statistically significant.

Several upregulated transcripts were also confirmed by real time PCR analysis. *Metallothionein D* (*MtnD*) was also found to be upregulated 4.17-fold according to the microarray and 5–fold by real time PCR in *Adar*^{5G1} flies (Figure 4.5). *MtnD* is one member of a family of four proteins involved in metal ion homeostasis. Interestingly, two other metallothionein genes were upregulated according to the microarray data, *MtnC* was 7.09-fold upregulated and *MtnB* was 1.48-fold upregulated. *MtnA* was not changed. Analysis of the evolution of the four genes indicates that *MtnA* is the ancestral gene, whereas *MtnB*, *-C*, and *-D* are arose from

several duplication events and are closely related (Egli *et al.*, 2006a). *Drosophila* mutants lacking all four *Mtn* genes display increased sensitivity to heavy metals and decreased life span (Egli *et al.*, 2006b). However, analysis of the contribution of individual *Mtn* genes to metal ion homeostasis revealed that *MtnA* and *-B* have the major role in metal ion homeostasis, whereas *MtnC* and *-D* play minor roles (Egli et al., 2006a). This preference was reflected in analysis of promoter responsiveness, as *MtnA* and *-B* were preferentially induced by copper and cadmium respectively. All four metallothionein genes are regulated by metal response elements in the promoter region which are activated by the metal-responsive transcription factor (MTF-1) (Zhang *et al.*, 2001). This indicates that *MtnC* and *-D* may have diverged in function from *MtnA* and *-B*, and may be induced in response to an as yet uncharacterised stress or cytotoxic agent. As metal ions can generate reactive oxygen species, metallothioneins also exert an antioxidant effect by chelating metal ions. Indeed overexpression of MTF-1 protects flies against heavy metal toxicity and oxidative stress (Bahadorani *et al.*, 2008). Hence, elevated levels of *Metallothionein* genes in $Adar^{5G1}$ flies may explain the increased resistance to oxidative stress observed in *Adar* mutant flies (Chen et al., 2004).

Expression of *CG30091* was approximately 30-fold upregulated in $Adar^{5G1}$ flies according to real time PCR, whereas the microarray data indicated it was 7.3–fold upregulated (Figure 4.5). The *CG30091* gene has not been characterised yet, although the molecular function attributed to it in Flybase is serine-type peptidase activity and therefore it is described as being involved in proteolysis.

Many primers were designed to analyse other transcripts that were altered according to the microarray data however despite giving product when tested in an RT-PCR reaction, they did not amplify transcripts in a linear fashion when analysed by real time PCR and therefore could not be used.

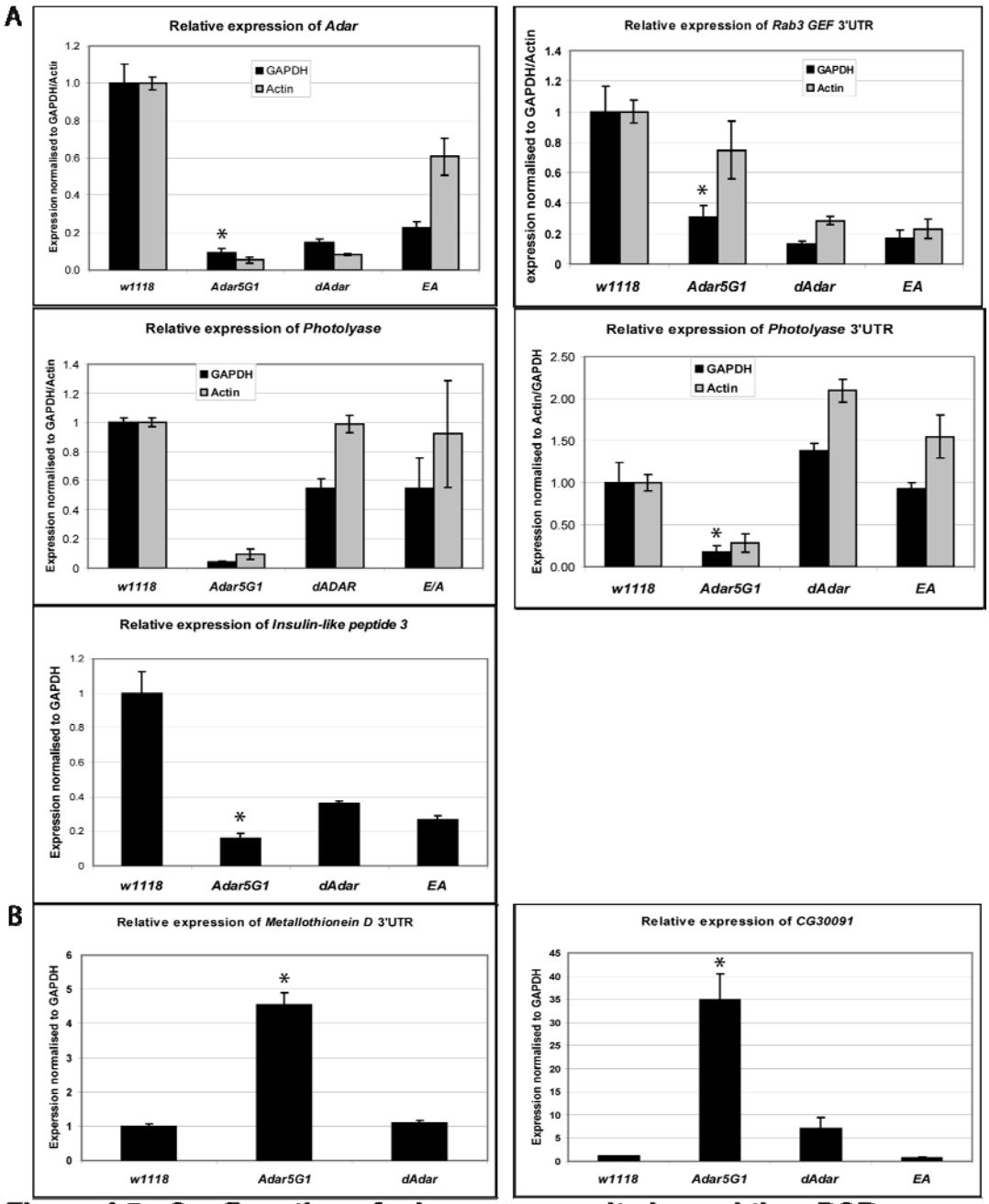


Figure 4.5: Confirmation of microarray results by real time PCR

A) Genes that were downregulated and B) Genes that were upregulated according to the microarray data were analysed by real time PCR. Initially transcripts were investigated using *Actin* (grey bars) and *GAPDH* (black bars) as reference genes, however broadly consistent data was obtained so further analysis was performed with *GAPDH*. Analysis of *Adar* transcript levels showed a 10-fold decrease in $Adar^{5G1}$ fly head, surprisingly the levels of *Adar* in the flies expressing *dAdar* in the cholinergic neurons was very low indicating low amounts of *Adar* transcript are required for editing of transcripts. Higher levels of transcript were seen in the inactive mutant (EA), reflecting a lack of toxicity of the inactive protein. *Rab3 GEF* levels were decreased in all genotypes compared to w^{1118}, consistent with the microarray data. *Photolyase* showed a 10-20-fold decrease in $Adar^{5G1}$ flies, with a similar profile accross the transcript as analysed by primers designed to the 3'UTR and the gene body. *Insulin-like peptide 3* was also decreased in $Adar^{5G1}$ flies, although the restoration in transgenic flies was modest here. *metallothionein D* and *CG30091* transcripts were upregulated in the absence of *Adar*. Data represents averaged results from 2 biological replicates and 2 technical replicates, error bars indicate +/- standard deviation. Where * indicates a $p < 0.05$ when comparing $Adar^{5G1}$ levels to w^{1118} levels using an unpaired t test.

Known substrates of ADAR enzymes show extensive secondary structure whereby stretches of dsRNA are interrupted by bulges and mismatches, which are important for ADAR-binding. Therefore a preliminary analysis of the predicted secondary structure of transcripts identified in the microarray analysis was performed to investigate whether these transcripts could be bound by ADAR. Secondary structure prediction software, RNAfold (University of Vienna) (Gruber *et al.*, 2008) was used to model the predicted structures of pre-mRNAs encoding *Ilp3*, *Phr*, and *CG30091* (Figure 4.6). These pre-mRNAs show extensive stem-loop structures with bulges and mismatches that could be bound by ADAR. However another transcript that was not altered according to the microarray data, *GAPDH2* also showed significant secondary structure suggesting ADAR may bind to many transcripts, both those altered on the microarray and those that remained unchanged.

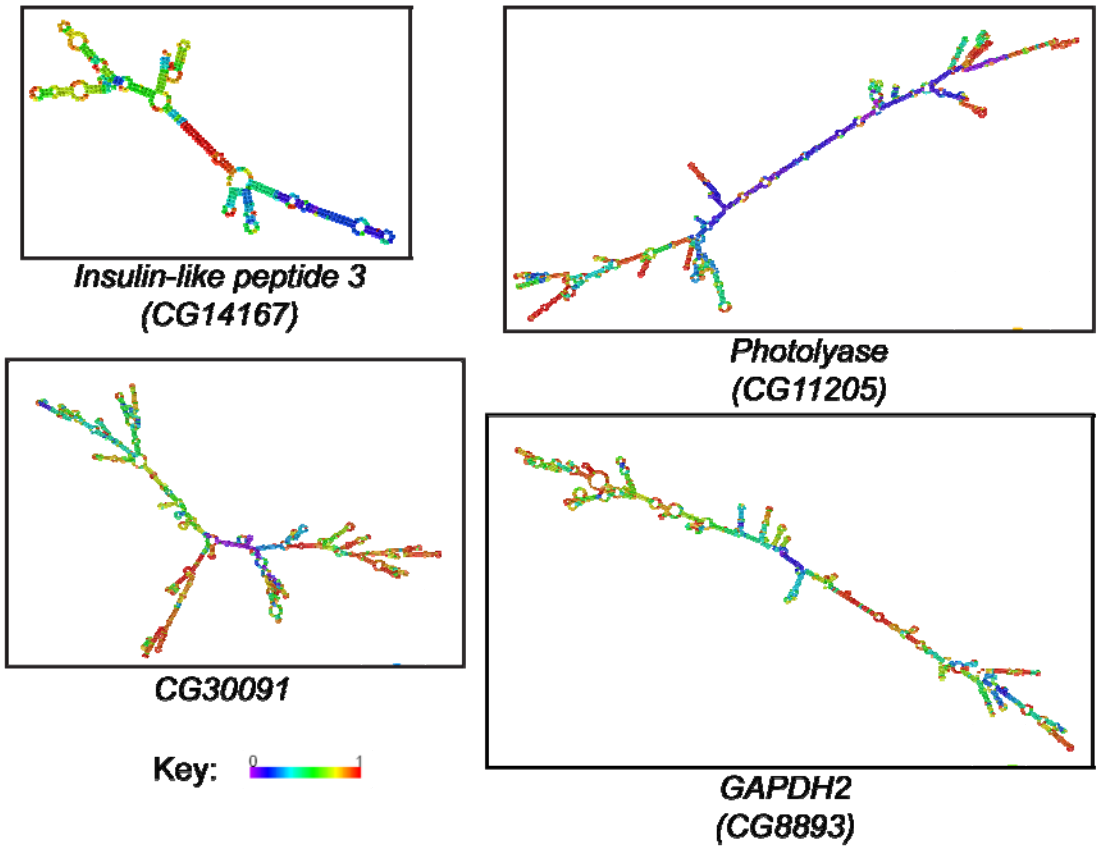


Figure 4.6: Predicted secondary structures of novel ADAR substrates
RNAfold secondary structure prediction software (University of Vienna; Gruber et al, 2008) was used to determine the secondary structures of candidate pre-mRNAs idenified from the microarray data as being altered in *Adar*^{5G1} flies and restored by expression of either active *Adar* or catalytically inactive *Adar EA* in the cholinergic neurons. *GAPDH2* was included as a negaitve control. Colours represent the base pair probabilities of the predicted structure.

When comparing w^{1118} and $Adar^{5G1}$ only a few known edited substrates of ADAR were present on the pair-wise microarray analysis list generated as part of the microarray service by Pawel Herzyk (data not shown). However these transcripts were not altered dramatically and they ranked low down on the rank profile list where the false discovery rate is increased. For example the *nicotinic acetylcholine receptor α-34E* gene (*nAcR-alpha34E*) was ranked number 343 showing a 1.66-fold increase in transcript levels in $Adar^{5G1}$ compared to w^{1118}, with a FDR of 4.2%. None were present in the lists generated by searching for specific patterns. However a lack of ADAR protein and site-specific editing may affect the levels of known substrates of ADAR through alterations in RNA processing or stability therefore a subset of edited transcripts were analysed by real time PCR. The genes analysed were: *nicotinic acetylcholine receptor α--34E (CG32975), nicotinic acetylcholine receptorα-30D (CG4128), nicotinic acetylcholine receptor β-64B (CG11348), resistance to dieldrin (CG10537), and Shaker (CG12348)*. All the transcripts analysed showed a similar profile across the four genotypes assessed (see Figure 4.7), however the changes were not dramatic (2-5-fold) when compared to the changes confirmed in *phr* (25-fold, Figure 4.6). *nAcR-alpha34E* was shown to be increased in $Adar^{5G1}$ flies according to the microarray data, yet the real time PCR indicates it is decreased. This disparity indicates the changes observed are probably not significant as they are subject to variability. Indeed previous reports have shown difficulty in confirming minor changes from microarray data, for this reason most analysis focuses on changes of greater than two-fold as these are more readily confirmed (Rajeevan *et al.*, 2001). Results shown in Figure 4.7 are the analysis of one biological replicate; however a separate replicate was performed with primers designed to the 3'UTRs of these transcripts to more accurately measure the same region of the transcript as the microarray. Both the gene body primers and the 3'UTR primers display the same profile across the four genotypes analysed (data shown for *Rdl* transcript).

The profile of expression of edited transcripts across the four genotypes analysed appears similar for the subset of edited transcripts analysed by real time PCR. The pattern is distinct from that seen with the novel substrates analysed by real time PCR

in Figure 4.6. This suggests that I have identified a novel class of transcripts that are regulated by expression of ADAR in a manner that is probably independent editing events.

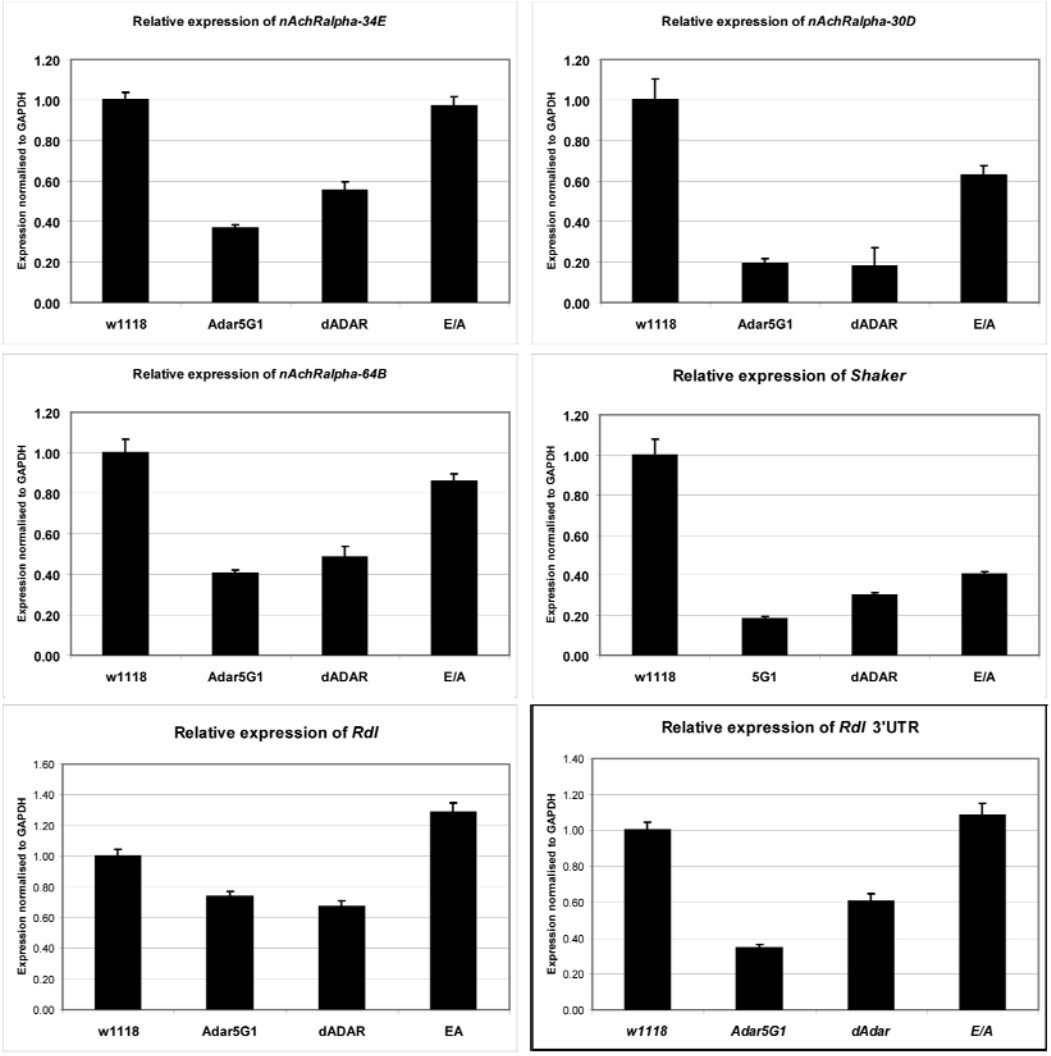


Figure 4.7: Real time PCR analysis of edited transcripts
As there is only one *Adar* gene in *Drosophila melanogaster* transcripts that show adenosine to inosine editing in the brain must be bound and edited by ADAR. Primers designed inside the gene and to the 3'UTR were used to analyse transcript levels in RNA extracted from w^{1118}, $Adar^{5G1}$, $Adar^{5G1}$:*Cha-GAL4>UAS-dAdar*, and $Adar^{5G1}$:*Cha-GAL4>UAS-dAdar E/A* fly heads. A consistent pattern of expression can be seen across the four genotypes with all transcripts showing lower levels in $Adar^{5G1}$ flies and higher expession in the flies expressing wild type or catalytically inactive *dAdar* in the cholinergic neurons. Data shown is the result of analysis of one biological replicate, however a separate biological replicate was analysed with primers in the 3'UTR and this confirmed these results. Only data from analysis with *Rdl* 3'UTR primers is shown.

One class of transcripts that were altered in *Adar*^{5G1} flies according to the microarray data were histone RNAs. Transcript levels changes were not large (maximum 2.4-fold lower in *Adar*^{5G1}) however histones 1, 2A, 2B, 3, and 4 were downregulated. Histone transcripts are not polyadenylated; instead they are regulated post-transcriptionally through a stem-loop structure in the 3'UTR (for a review see (Marzluff *et al.*, 2008)). As ADARs bind to stem-loop structures this could implicate ADAR in histone mRNA processing. However these transcripts were not restored by expression of active *Adar* or the catalytically inactive *Adar EA* transgenes in the cholinergic neurons, according to the microarray data. Further, an RNAi screen in *Drosophila* S2 cells to determine the genes involved in post-transcriptional regulation of histone mRNA processing did not identify *Adar* (Wagner *et al.*, 2007). Therefore the downregulation of histone transcripts in *Adar*^{5G1} flies may be indirect.

When Xia and colleagues (2005) attempted to identify new substrates for *Drosophila* ADAR using an antibody to inosine, they generated a list of 62 potential new substrates (Xia et al., 2005b). Further investigation of a subset of these transcripts showed that 7 out of 12 tested contained A-G changes indicative of editing. Comparison of the lists generated by searching the microarray data for patterns of expression with the list of 62 transcripts did not reveal any transcripts common to both lists. This further suggests I have identified a novel group of transcripts whose expression levels are regulated by *Adar* expression, independent of editing events.

In vitro analysis of ADAR binding to microarray-identified transcripts

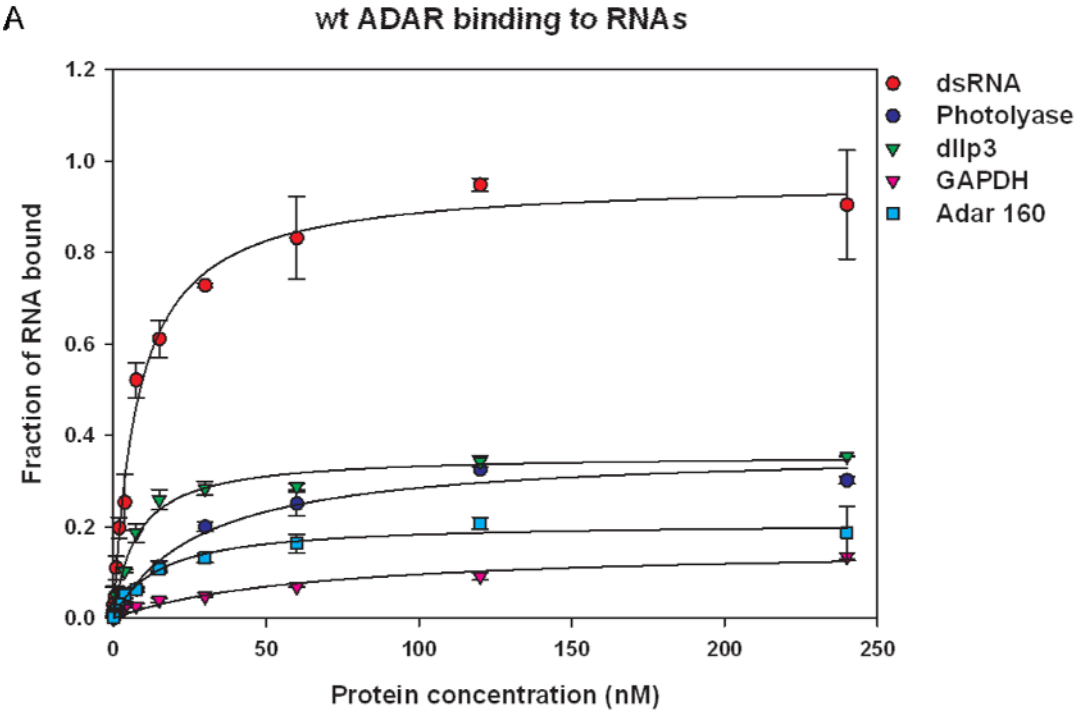
Pattern analysis of the microarray data generated two lists of transcripts that were potentially regulated by ADAR. However it was unclear whether ADAR can bind to these transcripts or whether the observed change in transcript levels was a secondary effect. To determine whether *A*DAR could interact with these transcripts several different methods were used. Binding of ADAR to candidate transcripts and controls was analysed *in vitro* using filter-binding assays. Briefly, recombinant epitope-tagged wild type or a catalytic site mutant of the ADAR protein was purified from *Pichia pastoris* using an existing protocol (Ring et al., 2004). As editing occurs co-transcriptionally before splicing takes place genomic DNA was cloned into T-easy (Promega) to allow transcription of pre-messenger RNAs for the candidate genes using either T7 or Sp6 polymerases (Ambion). The transcripts chosen for analysis were *Photolyase* and *Ilp3*, and controls *GAPDH* and *RP49* were included, the size of each pre-mRNA is listed in Table 4.4. The positive controls used were a 187bp fragment of the *Adar* transcript surrounding the self-editing site in exon 7 termed *Adar160* which has been shown to be edited to approximately 70% *in vitro*, and a perfect dsRNA fragment of 296bp prepared by *in vitro* transcription of ΔKP DNA and shown be extensively hyper-edited in a non-specific manner as described by Scadden & O'Connell (2005). Transcripts were incubated with varying concentrations of ADAR protein (0-240nM) for 10 minutes before reactions were applied to nitrocellulose filters. Filters were washed and allowed to air-dry, and bound RNA was determined by reading the radioactivity using a scintillation counter. Binding curves generated by plotting protein concentration against bound RNA were used to determine the dissociation constant (Kd) of the binding reactions (Figure 4.8).

Transcript	Length (bp)
Sp6-Adar160	**269**
ST7-Ilp3	**526**
Sp6-RP49	**1040**
Sp6-Cac	**1517**
T7-GAPDH	**1556**
Sp6-Phr	**1911**

Table 4.4: Size of *in vitro* pre-mRNA transcripts analysed by filter binding.
RNAs were *in vitro* transcribed from the promoters listed using either T7 or Sp6 polymerase (Ambion).

Filter binding analysis revealed that whilst recombinant ADAR bound with very high affinity to the perfect dsRNA as described previously (Gallo et al., 2003), binding to the other transcripts was with considerably lower affinity and maximal binding including the *Adar160* self editing site fragment. This result was unexpected as the *Adar160* fragment is edited to ~70% *in vitro* (Keegan et al., 2005), and therefore I was anticipating high affinity binding. This data suggests that binding affinity and editing efficiency are separate properties of ADAR. However, it has been suggested that different RNAs cannot be directly compared using the filter-binding method as the interaction between protein and RNA may differ for each transcript (Haynes, 1999). Further, the transcripts are of different lengths (see Table 4.4) and long transcripts may surround or coat the protein and allow it to pass through the filter. Therefore an alternative method was sought to validate these results.

A



B

	Kd (nM)	S.E.
dsRNA	8.0272	0.76
Adar160	14.5457	1.98
Ilp3	7.6389	0.89
Photolyase	28.9311	4.00
GAPDH	63.1074	16.32

Figure 4.8: Filter-binding analysis of wild type ADAR protein to different pre-mRNAs.

Varying amounts of wild type ADAR protein purified from *Pichia pastoris* were incubated with radiolabelled, *in vitro* transcribed pre-mRNA candidates identified in a microarray screen (*Photolyase* and *Ilp3*) or controls (positive controls dsRNA and *Adar160*, and negative control *GAPDH*). Binding reactions were applied to a nitrocellulose filter and washed. The amount of RNA bound by ADAR was calculated from scintillation counter readings of air-dried filters. Binding curves were plotted in SigmaPlot software package using the ligand binding equation y = (Bmax x [protein])/(Kd + [protein]), and Kd values shown in Table B were estimated from the curves, where S.E. indicates standard error. The perfect dsRNA positive control was bound by ADAR in a dose-sependent manner, reaching saturation by 240nM ADAR protein. The other transcripts analysed did not bind as well as the dsRNA control, however both Photolyase and *Ilp3* reached higher maximal binding than the *Adar160* control. The rate of binding (Kd) as estimated by the slope of the curve and shown in panel B indicates that ADAR binds the *Ilp3* transcript with higher affinity that the *Adar160* fragment known to be edited. Kd values for *Photolyase* and *GAPDH* are less reliable as these transcripts do not achieve maximal binding with the protein concentrations used therefore the Bmax is estimated.

The alternative method employed was to analyse ADAR RNA-binding by electrophoretic mobility shift assay (EMSA), as this method has been used to analyse binding of human ADAR2 to the GluR-B transcript (Ohman *et al.*, 2000). Binding of wild type dADAR (WT) to the *Adar160* self editing site was analysed.

An ADAR RNA-binding mutant (RRM) with similar mutations to those described in ADAR2 in Figure 4.11 was generated by site-directed mutagenesis. The proteins were expressed and purified in *Pichia pastoris* with the assistance of Jim Brindle using the protocol described by Ring and colleagues (Ring et al., 2004). The RRM mutant contains three amino acid changes such that lysine residues which are critical for RNA-binding are replaced with alanine or glutamine residues (KKxxK->EAxxA). ADAR1 and ADAR2 containing these mutations in their dsRBDs were described in (Valente and Nishikura, 2007). Mutation of these residues was designed to abolish RNA-binding without disrupting protein structure based on mutations made in *Drosophila Staufen* dsRBD3 (Ramos et al., 2000). However, the ADAR RRM protein was expressed at a lower level than wild type ADAR, see Appendix I. Purified protein was quantified using a Bradford assay and an equal amount was used for binding analysis.

EMSA analysis of binding of WT dADAR produced a shift in molecular weight with increasing protein concentration such that the entire RNA probe was bound at 400nM ADAR (Figure 4.9 A). The complex could also be super-shifted by incubation with an anti-FLAG antibody indicating the complex was formed by the FLAG-tagged ADAR. The complex could be competed by addition of increasing amounts of cold *Adar160* RNA, amounts analysed were 1ng, 10ng, 100ng, 500ng, and 1ug, which corresponds to 100-500,000-fold excess. *Adar160* transcript was transcribed *in vitro* in the presence of ^{32}P-GTP, and 10pM was used per reaction. Labelled and unlabelled RNAs were added to protein at the same time and reactions were incubated at 37^{o}C for 15 minutes before loading onto a native polyacrylamide gel. The EMSA was repeated twice and the bands quantified using ImageQuant software (GE Healthcare) then the data was plotted using Sigmaplot (Systat software Incorporated).

Competition analysis revealed the amount of cold competitor required to compete the reaction was between 10ng and 100ng. However the complex could not be disrupted by addition of a 1000-fold excess of yeast tRNA (5ug) indicating the interaction was specific. The equation listed in Figure 4.9 (C) was used to calculate the dissociation constant (Kd) of recombinant WT dADAR as 3.68 nM +/- 0.91. This is lower than the estimate based on the filter-binding assay which showed the Kd of wild type ADAR binding to *Adar160* RNA to be 14 nM +/-1.98. However, the results are consistent and taken together these results indicate wild type *Drosophila* ADAR has a dissociation constant within the range of 2.77-15.98 nM depending on the experimental conditions used.

In contrast the RNA binding mutant (dADAR RRM) did not show a significant shift in molecular weight until high amounts of protein were included (Figure 4.9 B). Interestingly the small shift that was observed could be disrupted by addition of 5ug yeast tRNA indicating the binding observed is non-specific. It has previously been reported that the ADAR1 deaminase domain is capable of binding to RNA in the absence or mutation of the dsRBDs (Herbert and Rich, 2001), which may explain the residual binding observed here, although this would imply weak non-specific binding is mediated through the deaminase domain as it can be competed by yeast tRNA. A catalytic site mutant dADAR EA protein showed similar binding to wild type and could not be competed off by yeast tRNA.

The dissociation constant of mammalian ADAR1 has been estimated at 0.11nM (Lai et al., 1995). This was calculated from filter-binding analysis with a perfect duplex dsRNA. Analysis of ADAR1 binding to a GluR-B fragment containing the Q/R site gave a Kd of 0.3nM (Lai *et al.*, 1997). ADAR1 does not preferentially edit the GluR-B Q/R site, however there is an intronic "hot-spot" site that ADAR1 edits which is included in the fragment analysed. The Kd of ADAR2 monomer binding to a GluR-B fragment as determined by EMSA was 1nM +/- 0.1 (Ohman et al., 2000). The binding of ADAR2 to fragment of the GluR-B transcript containing the Q/R site was determined to have a Kd of 0.2nM and binding to a fragment of the serotonin receptor gave a Kd of 0.5nM (Chen et al., 2000). In the same study the binding of

ADAR2 to perfect duplex dsRNA was estimated to be 0.2nM. ADAR3 which has no known substrates binds to the GluR-B fragment with a Kd of 0.4nM, the serotonin receptor fragment with a Kd of 0.6nM, and to perfect duplex dsRNA with a Kd of 0.5nM (Chen et al., 2000). This highlights the high affinity ADAR proteins have for RNA, regardless of whether deamination occurs. However, these reports indicate that there is variability between methods used to analyse binding and also between transcripts analysed.

Comprehensive analysis of the kinetics of ADAR2 binding to the GluR-B R/G site by EMSA showed that there was a progressive shift of RNA to a low molecular weight complex comprised of enzyme (E) and substrate (S) (Jaikaran et al., 2002), which were clearly visible as two distinct bands. With addition of more protein this shifted to a high molecular complex, comprising two monomers of enzyme and one molecule of substrate (ESE), this corresponds to the active editing complex. The dissociation constant for each complex was determined (ES = 4nM and ESE = 21nM). A similar stepwise shift was observed with ADAR2 and the GluR-B R/G site by Ohman et al, (2000), however the Kd for monomer binding was 1nM (Ohman et al., 2000). In comparison with the data presented here the GluR-B R/G fragment used for these experiments was 78bp long, whereas the *Adar160* fragment used in the experiments shown here was 269bp. A shorter RNA fragment would likely show a stepwise complex formation as observed for ADAR2 as it would be easier to resolve by EMSA. Importantly the filter binding analyses would not discriminate between monomer binding and dimer binding.

When comparing *Drosophila* ADAR to these mammalian ADARs it appears to have a lower affinity for the *Adar160* RNA, which is a known substrate. However the dissociation constants for ADAR2 determined by EMSA, are calculated for monomer binding and therefore direct comparisons with the data presented here for *Drosophila* ADAR cannot be made. The binding affinity is most similar to ADAR2, which is consistent with the high level of similarity (dADAR shares 67% identity with ADAR2 within the deaminase domain) shared between these proteins (Palladino et al., 2000a).

It has been shown that the dissociation constant for ADAR2 binding to an edited GluR-B R/G site fragment containing an inosine residue is lower than binding to the unedited transcript (Ohman et al., 2000). Therefore the fraction of edited RNA present in the sample may influence the RNA-binding; however it is not possible to control this variable without altering nucleotides within the substrate.

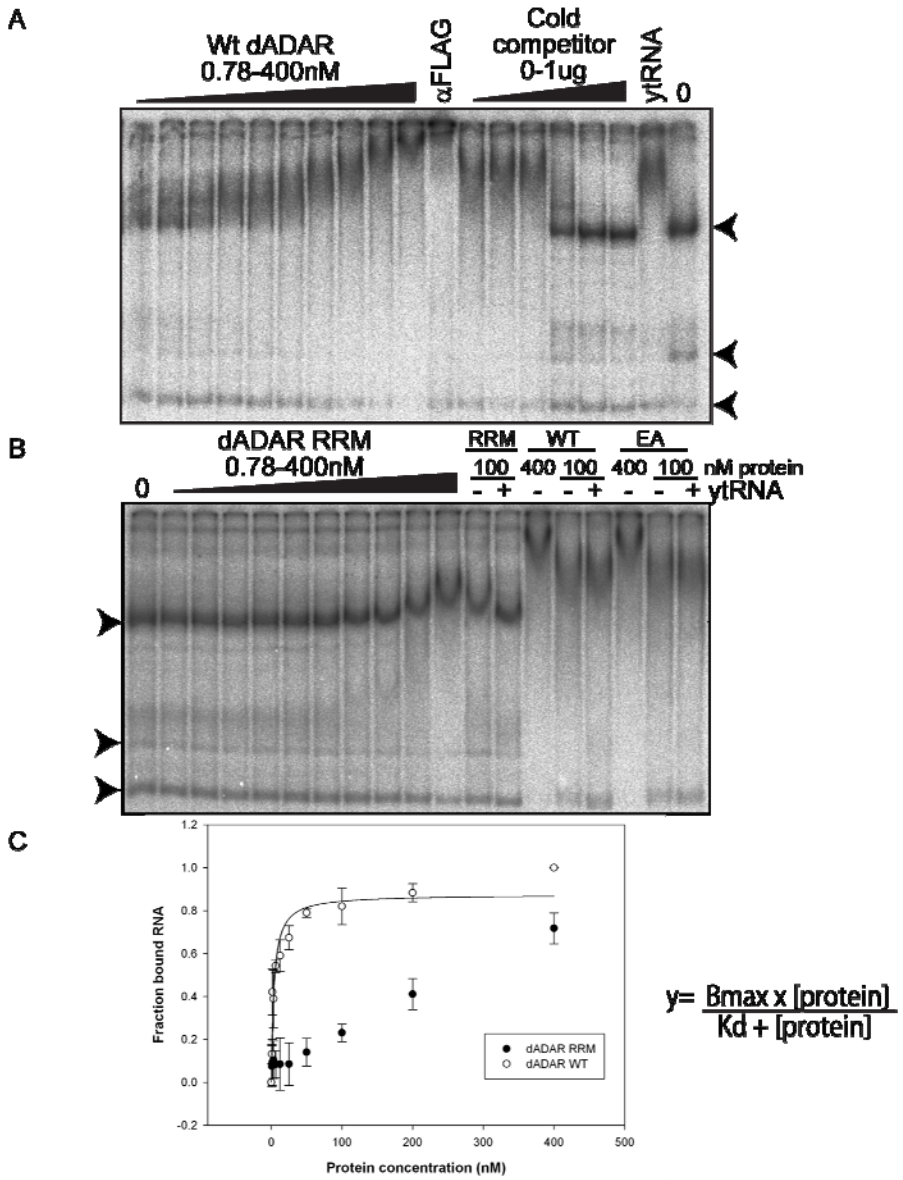


Figure 4.9: Binding of the dADAR RRM mutant to *Adar160* transcript is non-specific.

Binding of ADAR proteins to *Adar160* RNA was analysed by EMSA. Wild type *Drosophila* ADAR-FLIS6 purified from *Pichia pastoris* was incubated with ^{32}P-GTP radiolabelled *in vitro* transcribed *Adar160* RNA for 15 minutes at 37ºC. Reactions were stopped by loading onto a 5% polyacrylamide gel and electrophoresed for 7 hours at 4ºC to separate bound and unbound RNA. Free RNA is present as three bands indicated by arrowheads. Panel A shows the RNA shifts to a high molecular weight RNA-protein complex with increasing amounts of protein, such that all RNA is bound at 400nM ADAR protein. A supershift is observed with addition of 1ul of anti-FLAG antibody. This interaction can be competed with cold competitor *Adar160* RNA (0ng, 1ng, 10ng, 100ng, 500ng, 1ug), however it is not competed by addition of 5ug yeast tRNA (ytRNA) indicating the interaction is specific. In contrast the ADAR-FLIS6 RRM mutant has severely impaired binding, with only a small shift in molecular weight with 400nM protein. Interestingly this complex could be completely competed off with 5ug ytRNA indicating the weak interaction was non-specific. The catalytic mutant ADAR-FLIS6 EA protein has similar binding properties to wild type ADAR, and could not be competed with 5ug ytRNA. Panel C is a graphical interpretation of the binding data, the equation shown was used to fit a curve. Estimated Kd for dADAR WT is 3.68 nM +/- 0.91 , however Kd values could not be obtained for the ADAR RRM protein as the data does not fit the binding curve.

The binding of WT dADAR to *Adar160* was used to investigate the potential interaction to the transcripts identified as altered by microarray analysis. This was performed by analysing the ability of different transcripts to compete with the binding of dADAR to *Adar160* RNA. As shown in Figure 4.9 the interaction between WT dADAR can be partially competed by 10-100ng cold *Adar160* transcript, and completely competed with 500ng. However the interaction could not be competed with an excess of yeast tRNA, which has significant secondary structure. Therefore increasing concentrations of cold *in vitro* transcribed pre-mRNAs were incubated with 100nM dADAR protein and labelled *Adar160* RNA to determine if they were good competitors.

Figure 4.10 shows the EMSA gels and quantification of competition with *Adar160*, *Photolyase* (*Phr*), *Insulin-like peptide 3* (*Ilp3*), *RP49*, *GAPDH*, and *Cacophony* (*Cac*). The amounts of RNAs used were 0ng, 1ng, 10ng, 50ng, and 500ng, which correspond to 100-500,000-fold excess. These were incubated with 10pM radiolabelled *Adar160* RNA and 100nM WT dADAR protein for 15 minutes at 37^{o}C. The reaction was stopped by loading onto a 5% native polyacrylamide gel. Gels were electrophoresed for 7 hours at 4^{o}C, dried and exposed to a phosphorimager screen for 48 hours. Band intensity was quantified using ImageQuant software (GE Healthcare) and concentration of bound RNA was plotted against amount of competitor RNA. Cold competitor RNAs were analysed by amount rather than molar concentration so as not to assume prior knowledge of binding kinetics. However when differences in size of RNAs analysed was taken into account and the binding data was plotted against the molar concentration of the size of the RNA became apparent. This data suggests that ADAR binds multiple times to longer transcripts. Further experiments such as RNA footprinting are required to determine where the binding site for ADAR lies on these transcripts, and whether the interaction is stable. Although the positive control *Adar160* RNA has been shown to be edited at one site *in vitro* the predicted secondary structure for this molecule contains two stem-loop regions indicating ADAR could bind multiple times to this RNA.

It is unclear from this data whether any pre-mRNA tested would be bound by ADAR *in vitro*, as it would be difficult to find an RNA that does not show some secondary structure. Yet recombinant ADAR does not bind to yeast tRNA, so clearly some specificity exists.

Analysis of ADAR2 binding to a fragment of the GluR-B transcript has been used to analyse competition with other known substrates including a fragment of the serotonin receptor mRNA (Dawson et al., 2004). However, these experiments were done with fragments of the total transcript rather than the full pre-mRNA, and they specifically include the regions surrounding known editing sites as the quantification is based on editing frequency. This is the first investigation into binding of ADARs to transcripts that are not known to be substrates.

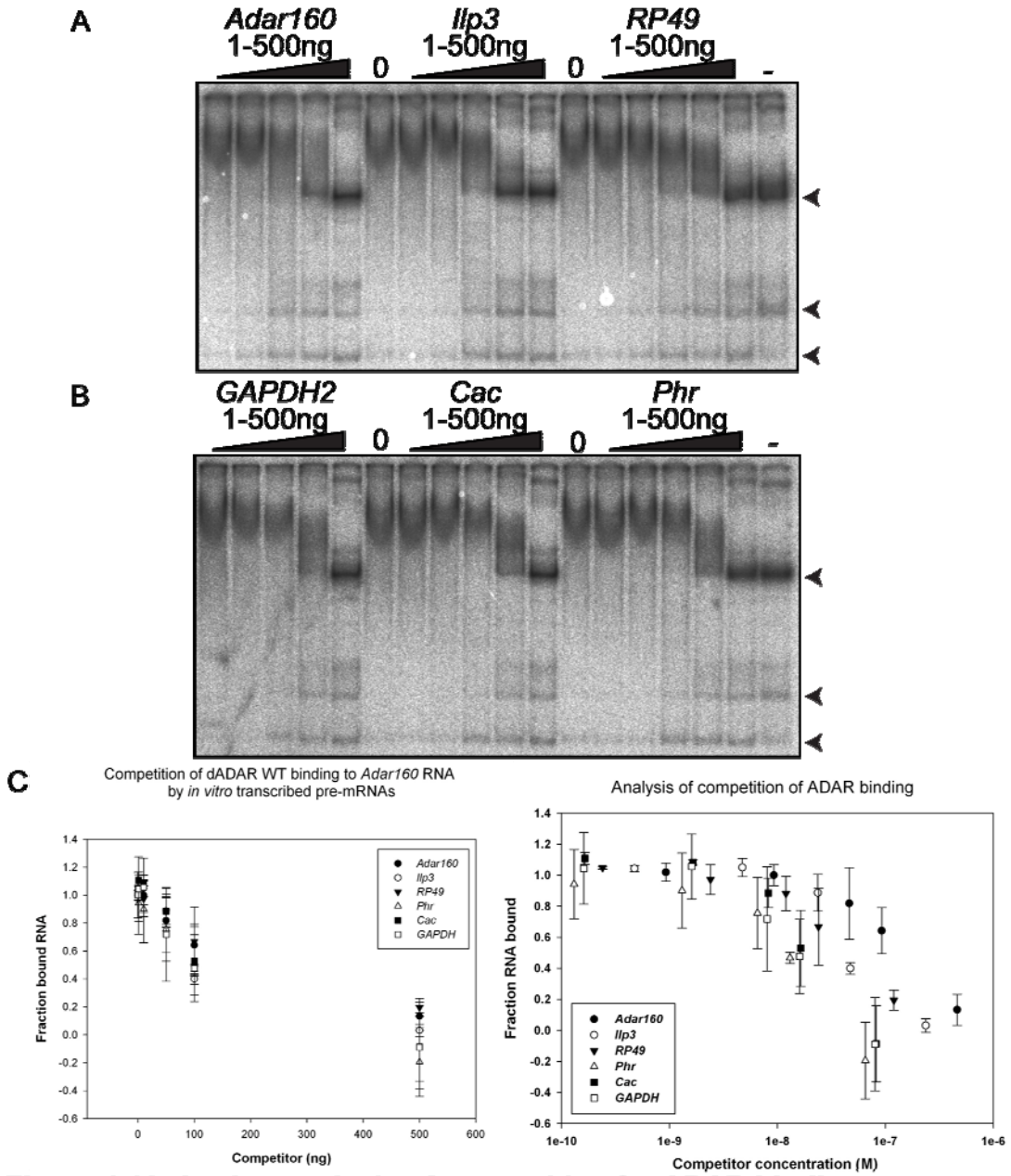


Figure 4.10: *In vitro* analysis of competition for ADAR binding

Binding of wild type ADAR protein to *Adar160* RNA was analysed in the presence of cold *in vitro* transcribed competitor pre-mRNAs. Reactions were incubated at 37°C for 15 minutes before loading onto a 5% polyacrylamide gel and electrophoresed for 7 hours at 4°C to separate bound and un-bound RNA complexes. Free RNA is indicated by arrowheads whilst RNA in complex with dADAR WT protein migrates slower. Amounts of cold competitor used were: 1ng, 10ng, 50ng, 100ng and 500 ng. At 500ng all RNAs tested were capable of competing the interaction between ADAR and the *Adar160* RNA. Quantification of bound and unbound RNA was performed using ImageQuant software and the results were graphed in C. Postive controls included were cold *Adar160* RNA and *Cacophany* (*Cac*), which are known to be bound and editied by ADAR *in vitro* and *in vivo*. pre-mRNAs tested were *Ilp3*, *Phr*, *GAPDH* and *RP49*. All RNAs showed a similar levels of competition for binding by ADAR, indicating ADAR binds to many substrates *in vitro*. However calculating the amount of RNA used in Molar concentration highlights that the difference in length (bp) of RNAs used determines the level of competition, which could indicate ADAR is capable of binding multiple times to the longer transcripts.

RNA-binding is an essential property of ADAR

In vitro data suggests that ADAR can bind to the transcripts that are altered in the *Adar*^{5G1} flies, along with transcripts that are not altered according to the microarray data. However to determine whether binding of ADAR to substrates is important for *in vivo* activity an ADAR mutant was made with mutations in several key residues in both of the dsRBDs. A construct was generated by site-directed mutagenesis of an existing construct containing human *Adar2*, this was injected injected into *Drosophila* oocytes and the resulting fly strain was crossed to a balancer line (Jim Brindle and Liam Keegan). However, as this mutant was specifically created for this project it was only available at the end of my PhD. The mutations used were initially described in (Valente and Nishikura, 2007), and they were specifically chosen to disrupt RNA-binding without perturbing the structure of the protein based on studies of other RNA-binding proteins including Staufen and PKR. The location of the mutations is shown in Figure 4.11 (A). The RNA-binding mutant termed *hAdar2 RRM* (KKxxK>EAxxA in dsRBD1 and 2) was expressed in the cholinergic neurons using the binary GAL4 system as previously described. Wild type human *Adar2* was used as a control as it has been shown to suppress neurodegeneration and locomotion defects (Leeanne McGurk, Thesis).

The flies expressing the *hAdar2 RRM* construct in the cholinergic neurons showed severe locomotion defects similar to the *Adar*^{5G1} strain (Figure 4.11 C). These flies showed holes in the brain characteristic of neurodegeneration, however they were smaller and fewer in number than those observed in *Adar*^{5G1}. 8μm sections were analysed by haematoxylin and eosin staining from a minimum of three flies, and although the holes are not as striking as those in *Adar*^{5G1} in the images in Figure 4.11 B, they were clearly visible under the microscope. The holes do not penetrate the thickness of the sections used, however thinner sections were lost during processing. This result indicates that RNA-binding is a crucial function for ADAR2. The presence of an ADAR protein may delay the onset of the neurodegenerative phenotype observed, however analysis of flies expressing the *hAdar2 RRM* at 30 days may show a more pronounced degeneration.

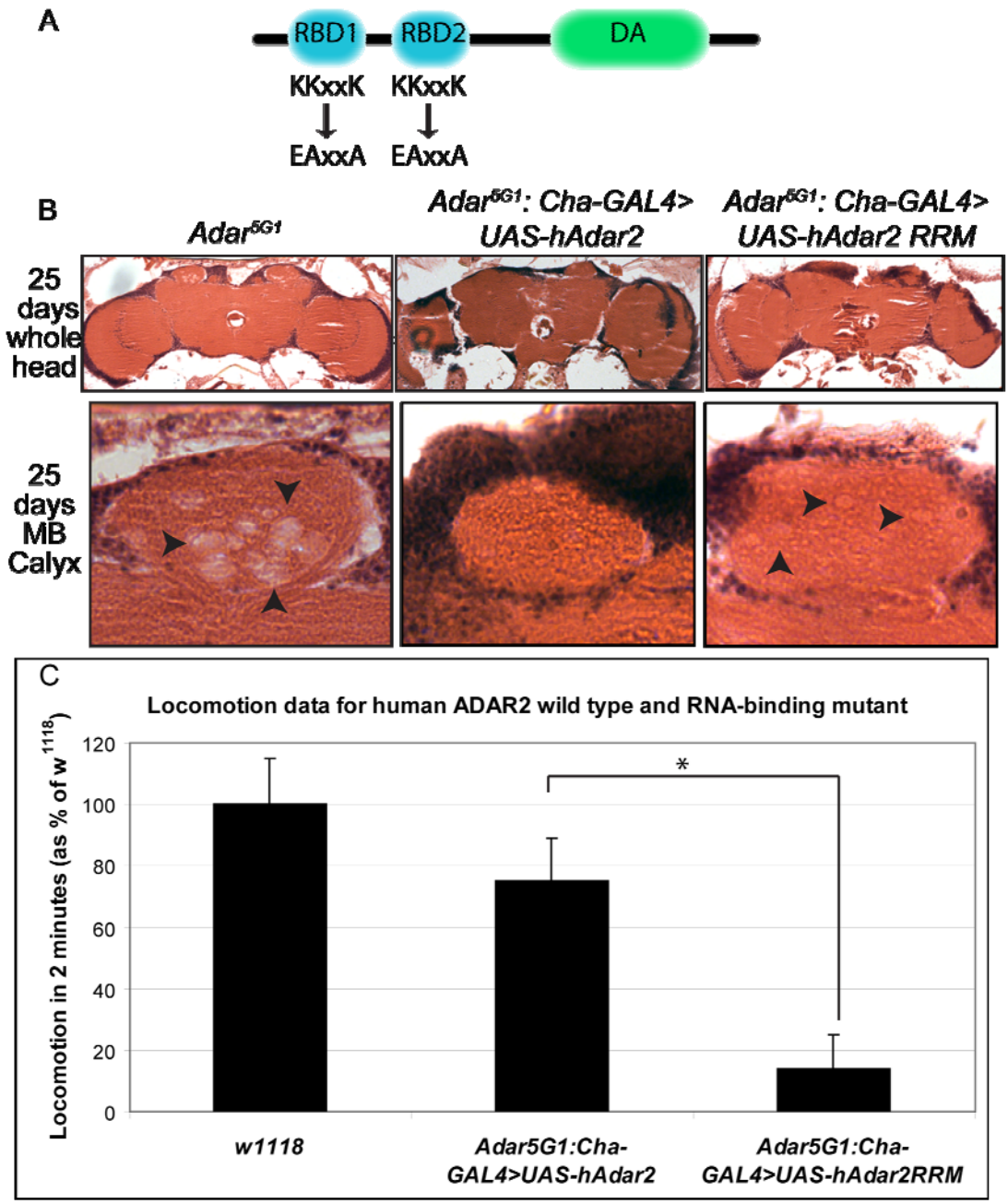


Figure 4.11: Human *Adar2* RNA-binding mutants show neurodegeneration and locomotion defects

A) A cartoon is shown depicting the location of the two KKxxK motifs in the *Adar* RNA-binding domains and the residues that they were mutated to. B) *Adar*-null flies expressing *hAdar2* in the cholinergic neurons show suppression of the neurodegenerative phenotype associated with the $Adar^{5G1}$ mutation. Holes in the brain are observed but they are fewer in number and smaller than those seen in $Adar^{5G1}$ flies. C) In a locomotion assay the hAdar2-expressing flies performed to a comparable level as w^{1118} flies. Flies expressing a *hAdar2 RRM* transgene with mutations in two KKxxK motifs located within the RNA-binding domain and crucial for RNA-binding were also analysed. These flies show numerous small holes in the brain (black arrowheads) and locomotion defects similar to $Adar^{5G1}$ flies, indicating RNA-binding is crucial for both locomotion activity and normal brain pathology. The difference in locomotion between flies expressing wild type *hAdar2* and mutant *hAdar2 RRM* is statistically significant when analysed by unpaired t test, $p<0.0001$ (indicated by asterisk) with a t value of 10.72.

Ex vivo analysis of Adar binding to microarray-identified transcripts

Biochemical analysis of transcripts identified as altered in *Adar*^{5G1} flies by microarray analysis demonstrated that ADAR can bind to these pre-mRNAs *in vit*ro. However ADAR also bound to transcripts that were unchanged by microarray analysis. To determine whether ADAR could preferentially bind to these transcripts in an *ex vivo* system an immunoprecipitation of ADAR was performed in *Drosophila* S2 cells followed by analysis of the bound transcripts. *Drosophila* S2 cells were transfected with a construct containing ADAR-FLIS6 under the control of a metallothionein promoter which is copper-inducible. Cells were transiently transfected with the construct or an empty vector control with Insect GeneJuice (Novagen) and treated for 24-hours with 0.6mM copper sulphate to induce expression of ADAR. Cells were harvested and protein concentration determined by Bradford assay. Equal amounts of protein were added to anti-FLAG agarose beads to immunoprecipitate ADAR, and following several washes, RNA was eluted by addition of TRIZOL reagent (Sigma). RNA was reverse transcribed with random hexamer primers (Promega) to ensure amplification of all transcripts bound and analysed by real time PCR. Transcripts were quantified with an aliquot of 10% of the input as determined by protein concentration.

Drosophila S2 cells are difficult to transfect with a maximal transfection efficiency of 20%; however immunocytochemical analysis of fixed cells with an anti-FLAG antibody revealed approximately 10% of cells were expressing ADAR (Figure 4.12 A). Non-transfected cells which were also treated with copper sulphate were used as a negative control. Non-specific cytoplasmic staining was observed in both control and transfected cells indicating the anti-FLAG antibody recognises endogenous epitopes in *Drosophila* S2 cells. However, cells transfected with FLAG-tagged ADAR showed nuclear staining which was absent from non-transfected controls and nuclear localisation has been observed for endogenous *Drosophila* ADAR protein in the brain (Leeanne McGurk, Thesis).

Immunoblot analysis revealed enrichment of ADAR protein following immunoprecipitation with an anti-FLAG antibody (Figure 4.12 B); however protein levels were too low to be detected by coomassie stain (data not shown). *Ataxin2* (*Atx2*) was used as a positive control as it is expressed in *Drosophila* S2 cells and it was identified as a substrate of ADAR by stapleton and colleagues (Stapleton et al., 2006). Investigation of transcripts immunoprecipitated revealed that the positive control *Atx2* was enriched by approximately 20% in the presence of ADAR, calculated relative to the 10% input fraction. The control immunoprecipitation which shows how much transcript can be bound by the beads showed a negligible amount of *Atx2*, approximately 1.2%. These values were used to calculate the enrichment of the *Atx2* transcript when ADAR was present using the following equation: transcript immunorecipitated in the presence of ADAR divided by transcript immunoprecipitated with no ADAR expression. The enrichment of the *Atx2* control was approximately 16%. The other transcripts investigated showed lower levels of transcript immunoprecipitated by ADAR; however they also showed lower levels of background binding to beads. Calculating the enrichment revealed that *Photolyase*, *Actin* and *GAPDH* all showed a similar level of enrichment of approximately 8-fold, which was half that of the positive control *Atx2*.

These results indicate that ADAR does not preferentially bind to the *photolyase* transcript *in vivo* as the level of enrichment did not differ from the negative controls of *Actin* and *GAPDH*. Enrichment of binding was observed for the positive control *Atx2*. However, using non-transfected S2 cells as a control was not ideal. Following site-directed mutagenesis to disrupt key residues in the RNA-binding domain of *Drosophila* ADAR an ADAR RRM mutant was generated for use as a negative control. However immunoblot analysis revealed this protein was expressed at approximately half the level of the wild type protein (see Appendix I and Figure AI.2) and immunoprecipitated amounts also differed (data not shown). Despite many attempts to optimise this experiment I was unable to repeat it with more appropriate controls. Therefore the data suggests that the binding of ADAR to transcripts identified by microarray analysis that was demonstrated *in vitro* is not observed in a cell culture model.

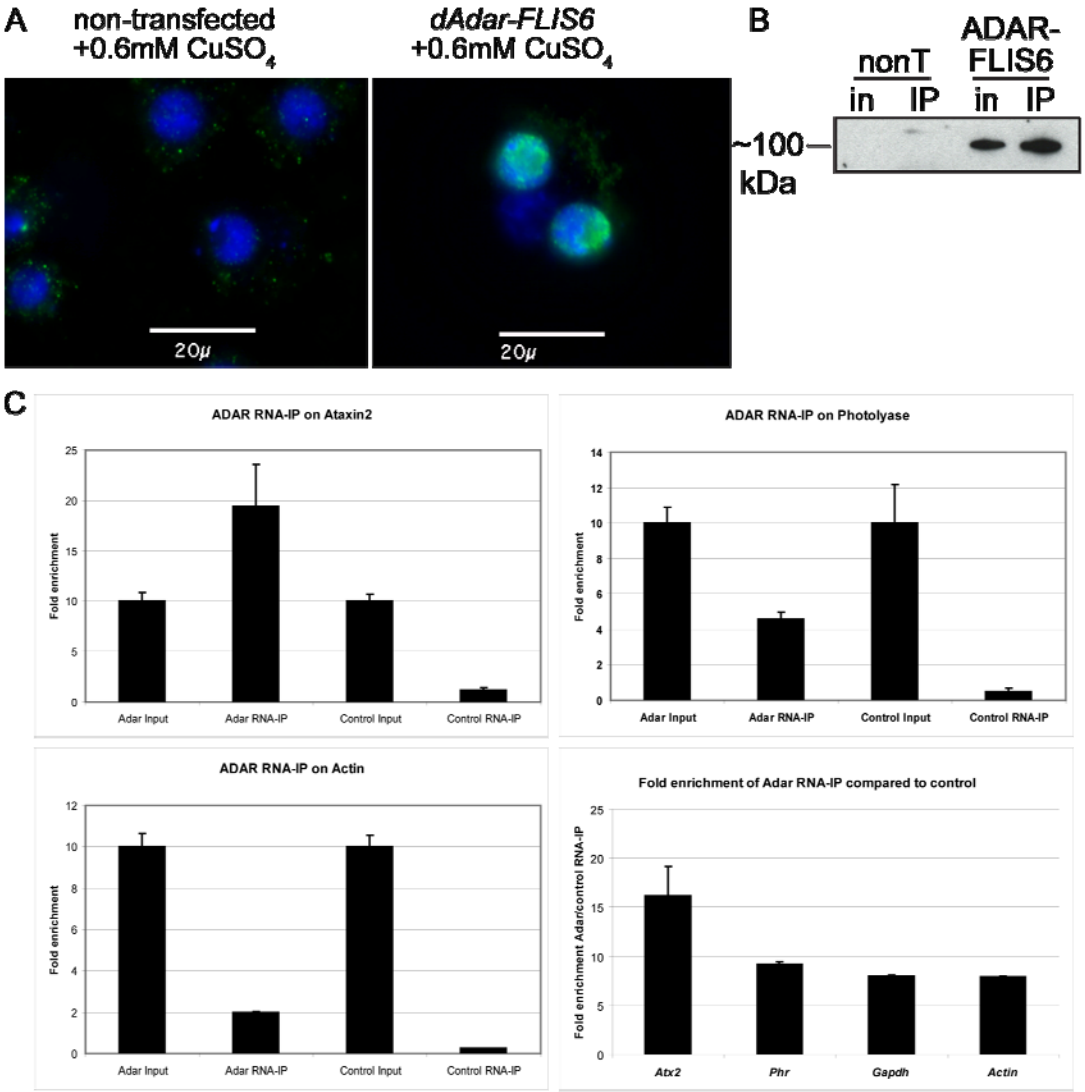


Figure 4.12: Immunoprecipitation of ADAR-RNA complexes from *Drosophila* S2 cells

Drosophila Schneiders S2 cells were transiently transfected with a vector expressing ADAR-FLIS6 under the control of a inducible metallothionein promoter. Following 24 hours treatement with 0.6mM copper sulphate cells were fixed and ADAR protein was visualised with an anti-FLAG primary antibody (Sigma) and a FITC conjgated secondary antibody. Panel A shows approximately 10% of transfected cells were expressing ADAR following induction, and ADAR protein showed nuclear localisation. Nuclei are stained with DAPI. Cells were harvested and ADAR was immunoprecipitated with anti-FLAG agarose beads. B) Protein aliquots were analysed by immunoblot which revealed enrichment of ADAR protein following immunoprecipitation. C) RNA was extracted with TRIZOL (Sigma), reverse transcribed and analysed by real time PCR. *Ataxin2* was used as a positive control as it is known to be edited and is also expressed in S2 cells. Amount of transcript was calculated according to input RNA which was used to generate a standard curve. Fold enrichment was calculated as amount of transcript present when ADAR was expressed divided by amount of transcript present in the control immunoprecipitation. *Ataxin2* showed approximately 15-fold enrichment in cells expressing *Adar* compared to control. *Photolyase*, *Gapdh* and *Actin* all showed similar levels of binding, with approximately 8-fold enrichment in cells expressing ADAR when compared to control cells. Where error bars indicate +/- 1 standard deviation.

Editing of novel Adar substrates

A surprising result from the microarray analysis was that the levels of known substrates of ADAR were not dramatically altered. However, recent high-throughput sequencing of candidate transcripts revealed that many more mammalian RNAs are edited than was previously thought (Li et al., 2009). The *photolyase* and *Ilp3* transcripts investigated in this thesis are not known to be substrates of ADAR, however further analysis was required to determine whether they are edited *in vivo*.

Photolyase and *Ilp3* were amplified from w^{1118} head RNA and sequenced to determine whether they were edited *in vivo*. No editing was observed in 70 sequence reads from *Ilp3* transcript. The *Photolyase* transcript is longer and was less abundant therefore sequencing was more difficult and data was only obtained for 12 full length transcripts, which showed no evidence of editing. However, editing may occur at a specific developmental stage or in response to a specific stimulus. Therefore, to determine whether these transcripts can be edited by *Drosophila* ADAR an *in vitro* editing assay was performed. Conditions were optimised using the *Adar* self-editing site fragment *Adar160,* and recombinant ADAR protein. Figure 4.13 (A) shows the results of sequencing a mixed population of *Adar160* cDNA following incubation with recombinant dADAR protein. A mixed peak indicates the site-specific editing event. No editing was observed with mutant dADAR RRM or dADAR EA, indicating both binding and catalysis are required for editing.

In vitro transcribed *Ilp3*, *Phr*, or *RP49* were incubated with 50nM recombinant ADAR for 2.5 hours at 37^{o}C. RNA was extracted with TRIZOL (Sigma), reverse transcribed and cloned into the T-easy vector (Promega) for sequence analysis with primers specific for the T7 and Sp6 promoter regions. Sequence analysis of 78 transcripts of *Drosophila Ilp3* revealed no clear "hotspot" site for editing. Single point mutations observed were spread randomly throughout the clones analysed and no significant difference was observed for A-to-G mutations over G-to-A mutations. However one clone exhibited A-to-G mutations at multiple residues within a short

60bp fragment. This is reminiscent of hyper-editing where up to 50% of adenosines are edited to inosines.

Similarly, hyper-editing was observed in one clone of the *photolyase* transcript out of approximately 40 clones analysed and other point mutations were spread randomly throughout the sequence indicating no clear site-specific editing (see Figure 4.13 B). *Photolyase* was divided into three regions for sequence analysis as it is too large to sequence in one read. Indeed, when the 60bp fragment containing the changes was analysed further, in *Ilp3* only 5 other adenosines were present indicating 58% editing for one transcript, whereas in *Phr* there were 7 other adenosine residues giving 50% editing in one transcript. This hyper-editing occurred within the coding region and at the 3' end of each transcript. This hyper-editing was not observed in 41 clones sequenced from a control transcript *RP49*. The point mutations observed are likely to be PCR or sequencing artefacts.

One example of hyper-editing has been reported *in vivo* in the *Drosophila 4f-rnp* transcript where 263 A-to-G mutations were observed in a 4kb mRNA (Petschek et al., 1996). However, despite extensive sequence analysis this hyper-editing was only observed in one cloned *4f-rnp* cDNA (Peters et al., 2003). The dsRNA duplex containing *4f-rnp* was formed with a neighbouring transcript *sas-10*, which showed developmentally regulated read-through of a transcription termination signal. It is intriguing that I have observed the same phenomena occurring *in vitro* through intramolecular dsRNA formation; however it is unclear what role the biological significance of hyper-editing is.

This type of hyper-editing like that described here (and seen in Figure 4.13) has been observed for mammalian ADARs when analysing perfect duplex RNAs (Scadden and O'Connell, 2005). At low protein concentration the ADAR protein appears to bind to one transcript and edit it multiple times, producing an all or nothing effect where one transcript shows high levels of editing whilst others remain unedited (Mary O'Connell, unpublished observation). This indicates that more editing may be observed with a higher concentration of protein, however it is unlikely to be site-

specific editing as the control *Adar160* RNA fragment was efficiently edited in a site-specific manner under the same conditions.

Hyper-editing has also been observed for ADAR1 when analysing the GluR-B R/G editing site *in vitro*, but was rarely seen with the ADAR2 enzyme which showed much higher site selectivity (Kallman *et al.*, 2003). In this study no evidence of promiscuous editing was observed *in vivo* indicating this may be an artefact of the *in vitro* editing assay. Indeed, altering the substrate conformation by introducing point mutations was sufficient to reduce site-specific editing of the GluR-B R/G site by ADAR2 from 80% to 30% indicating for ADAR2 the catalytic activity is highly determined by the local structure of the substrate (Kallman et al., 2003). This indicates that *in vitro* the dADAR protein may behave more similarly to mammalian ADAR1, showing less substrate specificity as it has to edit a wider range of *Drosophila* transcripts.

Previously *in vitro* editing assays have been employed to analyse the structural requirements and specificity of site-specific editing on known substrates of ADARs. As such, specific and promiscuous editing can easily be defined. The sequence changes observed in *phr* and *Ilp3* results are indicative of promiscuous editing, which confirms that these transcripts form secondary structures that ADAR can bind to *in vitro*, although I cannot conclude from these experiments whether *Ilp3* or *Phr* are substrates of ADAR *in vivo*. *In vitro* editing assays are usually performed on known substrates to analyse the kinetics of the interaction rather than to identify novel substrates. However as the hyper-editing is not observed with the control *RP49* RNA this suggests that dADAR has an affinity to *phr* and *Ilp3*.

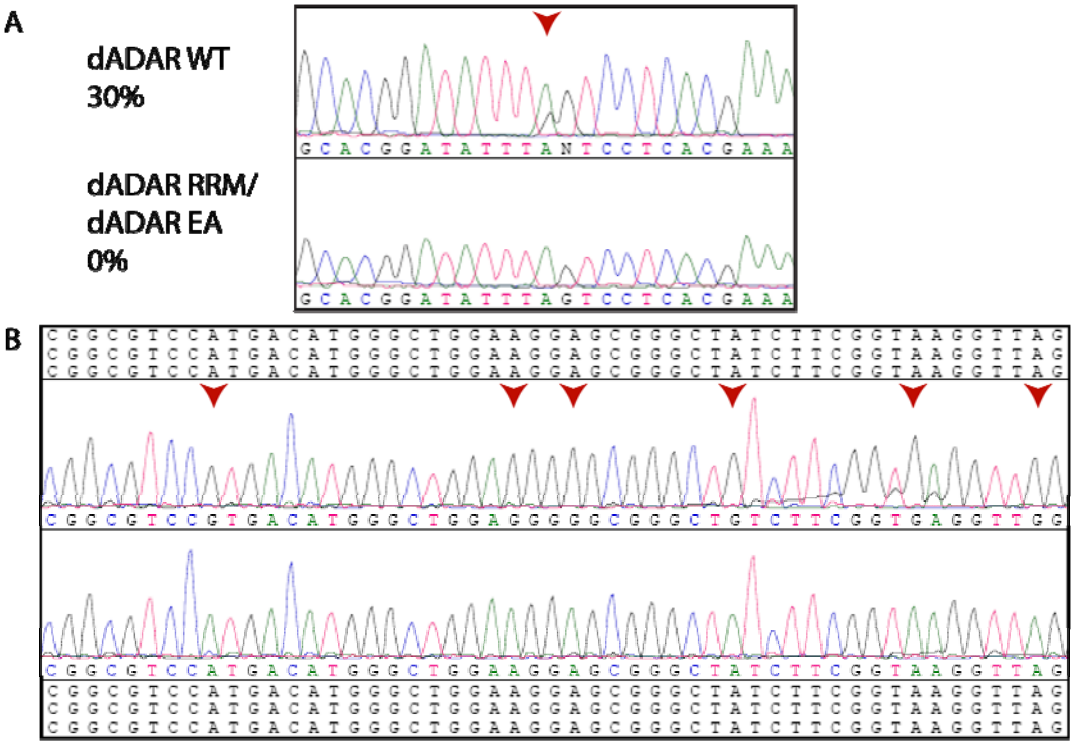


Figure 4.13: *In vitro* editing assay results.
A) *In vitro* transcribed RNAs were incubated with recombinant *Drosophila Adar* at 37°C for 3 hours. RNA was isolated, reverse transcribed and cloned into T-easy (Promega) for sequence analysis. Initial conditions were optimised for the *Adar160* fragment containing the *dAdar* self editing site. As the location of this site is known a mixed population of cDNAs were directly sequenced revealing two dye peaks at the editing site. Calculating the proportion of guanosine present compared to the total adenosine and guanosine gives an approximate value of 30% editing in the presence of WT dADAR. No editing was observed with mutants dADAR RRM or dADAR EA. B) For the other transcripts analysed individual clones were sequenced. For both *Ilp3* and *Phr* transcripts one clone was identified with multiple sequential editing events, *Phr* 3' region is shown in panel B, where 7 adenosines have been mutated to guanosines following incubation with WT dADAR, 6 are indicated here by red arrowheads. This was only observed in one sequence of 45 analysed in this region and was not seen in other regions of the *Phr* transcript. Wild type sequence is shown below for comparison.

Analysis of transcript levels in Dicer-2 mutant flies

The processes of RNA interference and RNA editing both require dsRNA formation. In C.elegans defective chemotaxis observed in *adar*-null (*adr-1/adr-2*) worms can be rescued by mutations in the RNAi pathway components *rde-1* and *rde-4* (*C.elegans* homologs of AGO-1 and TRBP-1) indicating the phenotype associated with loss of *adar* in worms is due to aberrant RNAi (Tonkin and Bass, 2003). Therefore the observed changes in transcript levels in the absence of *Adar* could be explained by an

inhibition or antagonism of the RNAi pathway by ADAR which normally occurs in w^{1118} flies, but which is lost in $Adar^{5G1}$ flies. A loss of inhibition of RNAi would result in a decreased transcript level in $Adar^{5G1}$ flies. Therefore, to determine whether transcript changes observed in $Adar^{5G1}$ flies are due to alterations in the RNAi pathway candidate transcripts were analysed by real time PCR in flies with defective RNAi due to mutation of the *Dicer-2* gene. If the hypothesis is true and RNAi is affecting the candidate transcripts then we would expect to see elevated transcript levels in *Dicer-2* mutant flies where siRNA production is impaired.

Drosophila melanogaster contains two dicer genes (*dcr-1* and *dcr-2*), which both contain two ribonuclease domains and an RNA-binding domain but differ outside these regions (Lee *et al.*, 2004). Depletion of both genes severely impairs production of short RNAs and RNA interference in S2 cells (Bernstein *et al.*, 2001). *Dicer-1* contains a PAZ domain but does not contain a functional helicase domain normally found in DICER family members. *dcr-1* is preferentially involved in the miRNA production pathway and as such it is an essential gene. *dcr-2* however is not essential and functions in the production of siRNA. DICER-2 contains a canonical DExH-type ATP-dependent RNA helicase domain, but does not contain a PAZ domain (a cartoon depiction of DCR-2 is shown in Figure 4.14 A). Point mutations in *dcr-2* generated by EMS mutagenesis show impaired siRNA generation, however low levels were present indicating *dcr-1* may also be able to compensate for *dcr-2* in the siRNA pathway (Lee et al., 2004). Two of the mutations analysed were nonsense mutations which resulted in truncated proteins lacking the ribonuclease domain, whilst others were missense mutations in the helicase domain. Interestingly all the mutants have similar low levels of siRNA, indicating the nonsense mutations in the helicase domain have the same phenotype as the ribonuclease-null mutants. Further investigation revealed a separate role for *Dicer-2* downstream of siRNA production as response to a synthetic siRNA is also impaired in these mutants.

Rab3 GEF was downregulated in $Adar^{5G1}$ flies according to the microarray and confirmed by real time PCR in Figure 4.5. The 3'UTR region of *Rab3 GEF* overlaps with the neighbouring transcript *Cyp4s3*, which was also downregulated 2.45-fold in

Adar^{5G1} according to the microarray data. This confirms that both of these transcripts are expressed in the head. If transcripts were produced from these convergent genes in the same cell then perfect duplex dsRNA may form, providing a substrate for ADAR or for the RNAi pathway. Interestingly *Rab3 GEF* has been shown to be edited by *Adar* in the 3'UTR region indicating dsRNA is formed in this region and ADAR binds to it. *Photolyase* was also investigated as the 3' region of the gene shows a short (<200bp) overlap with a neighbouring gene which is transcribed from the opposite strand, *pros*α*6,* and both of these transcripts were downregulated in *Adar*^{5G1} flies. Further, as described earlier the *photolyase* 3' region has been duplicated; therefore siRNAs produced from one of the *photolyase* transcripts may target the other transcript for RNAi. To determine whether RNA interference was impaired in the *Adar*-null flies I performed real time PCR on candidate transcripts *photolyase* and *Rab3 GEF* in *dcr-2* mutant flies. Three different *dcr-2* mutants were used (G31R, R416X and A500V) which contain nonsense or missense mutations in the helicase domain and have been shown to be severely impaired in siRNA production (Lee et al., 2004), (see Figure 4.14 panel A for cartoon depicting the location of the mutations within the *dicer-2* gene).

Two different primer sets were used to investigate these transcripts, one in the body of the gene and one at the 3'UTR where the Affymetrix probes are located. The reason for this is to determine which *photolyase* transcript is being analysed as the 3'UTR primer pair will detect transcripts from both genes (*CG11205* and *CG18853*) while the primers in the gene body are specific to gene *CG11205*. The Affymetrix probes used for microarray quantification are located in the 3'UTR and detect transcripts from both genes. The data presented shows the result of one experiment where real time analysis was performed in duplicate; however the three different *dcr-2* mutants all behaved in a similar way.

Transcript levels were normalised to *GAPDH* levels and expression is shown relative to *w*^{1118} levels. Comparison of the levels of *Rab3 GEF* in all fly genotypes indicates a similar profile across the transcript with levels low in *Adar*-null flies and not showing recovery with re-expression of active or inactive *Adar*, which confirms the

microarray data. This expression pattern is similar to the pattern observed for other known substrates of ADAR (compare with Figure 4.7). However, in the three *dcr-2* mutant fly strains levels of *Rab3 GEF* are slightly lower than w^{1118} (ranging from 0 to 2-fold difference, see Figure 4.14 B) but this not as severe as seen in the *Adar*-null flies (~2.5-fold). If *Adar* were acting to antagonise siRNAs produced from the *Rab3 GEF* locus we would expect the *dcr-2* mutants to show a similar or stronger phenotype as siRNA production is severely impaired in these mutants.

Analysis of *Photolyase* revealed that while transcript levels within the gene and at the 3'UTR are similar in *Adar* mutant flies, they show dramatic differences in *dcr-2* mutant flies (see Figure 4.14 B). The primers within the *photolyase* gene which are specific for gene *CG11205* show approximately 25-fold reduction in transcript level in $Adar^{5G1}$ flies, and a similar reduction (14 to 33-fold) in all three *dcr-2* mutant fly strains indicating a similar pathway is perturbed in both flies. However analysis of the 3'UTR region (shared between *CG11205* and *CG18853*) showed a similar decrease in $Adar^{5G1}$ flies of 10-fold but this was not observed in the *dcr-2* mutant flies. The *dcr-2* mutants showed levels similar to or higher than w^{1118} levels. This indicates that *dcr-2*-null flies show expression of transcripts from the *CG18853* locus that are seen in w^{1118} but are not observed in *Adar*-null flies.

Taken together these results indicate that *Adar* and *dcr-2* are not functioning in the same pathway as transcript levels are not consistent between mutants. If ADAR were antagonising the RNAi pathway for these specific transcripts we would expect lower levels of transcript in the *Adar*-null flies where antagonism is relieved which is what we observe. However we would also expect levels of these transcripts to rise in the absence of *dcr-2* which is not observed, indicating these transcripts are not altered in *Adar*-null flies due to antagonism of the siRNA pathway. As mutations in *dcr-1* are lethal it is not possible to examine in this way whether the miRNA pathway is altered in *Adar*-null flies.

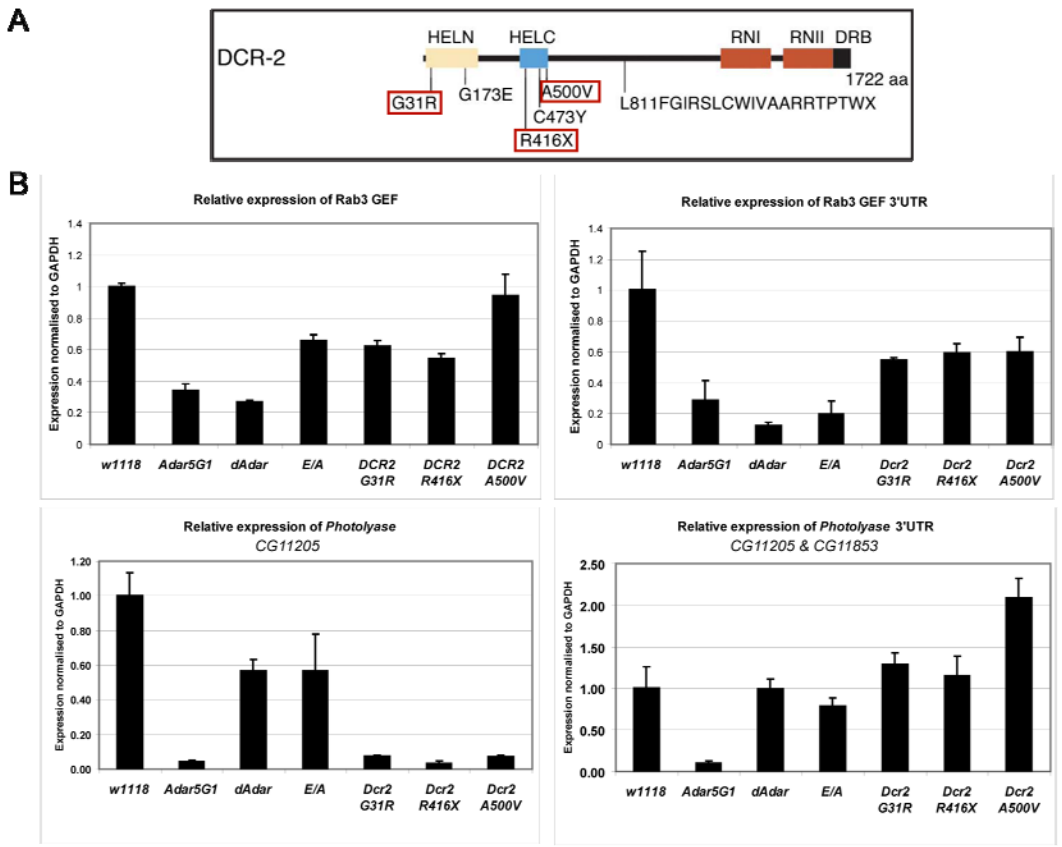


Figure 4.14: Analysis of transcript levels in *Dicer*-2 mutant flies.
To determine whether antagonism of the RNAi pathway was responsible for the altered levels of *Photolyase* and *Rab3 GEF* transcripts in *Adar*-null flies real time PCR was performed on RNA from flies with mutations in *Dicer*-2 (*dcr-2*). A) Mutations are indicated on a cartoon of Dicer-2 protein with the 3 used here highlighted in red boxes. Figure modified from Lee *et al*, (2004). B) Real time PCR results shown for *Photolyase* and *Rab3 GEF*. Both transcripts were analysed with primers in the gene body and a separate pair in the 3'UTR. This was to determine if levels were consistent across the transcript and also as the gene body primers in *Photolyase* are specific to the *CG11205* transcript, whereas the 3'UTR primers will detect both *CG11205* and *CG18853*. Real time PCR analysis indicates that the *RAB3 GEF* transcript is low in $Adar^{5G1}$ flies and is not rescued by expression of wild type *Adar* in the cholinergic neurons, consistent with the microarray data. The catalytic site mutant *Adar EA* shows higher levels of *RAB3 GEF* transcript, but they remain still lower than w^{1118}, and there appears to be a difference in transcript level between the 5' and 3' regions although this is difficult to explain. In the *dcr-2* mutant flies *RAB3 GEF* levels are consistently lower than w^{1118}, if this transcript were regulated by RNAi we would expect a higher amount of transcript in *Dicer-2* mutant flies. Photolyase levels are between 10 and 25-fold lower in $Adar^{5G1}$ flies with both primer pairs, which is consistent with the microarray data. Expression is rescued to levels similar to w^{1118} in flies expressing either the active or the inactive *Adar* in the cholinergic neurons. However the results form the *dcr-2* mutant flies show a difference between the primer pairs indicating the *CG11205* transcript is very low in the *dcr-2* mutants and the *CG18853* transcript is comparable to w^{1118}. If RNAi were acting on the *Photolyase* transcript(s) we would expect transcript levels to increase in the *dcr-2* mutant flies as the production of siRNAs is lost in these flies. Transcript levels are normalised to *GAPDH* and displayed relative to w^{1118} levels. The three *dcr-2* mutants displayed comparable levels of the transcripts analysed.

Does RNA oxidation influence transcript levels in Adar-null flies?

There has been a lot of recent interest in the role of nucleic acid oxidation in neurodegenerative disease, specifically RNA as is more vulnerable to oxidation than DNA. In Alzheimer's disease (AD) and Parkinson's disease (PD) elevated levels of RNA oxidation was shown to precede neuronal cell death and the presence of pathological inclusions indicating this is a primary pathological event (Nunomura *et al.*, 2009). Interestingly, the oxidised RNA was present only in the distinct subset of cells that would go on to develop pathological inclusions, for example dopaminergic neurons in the substantia nigra of patients with Parkinson's disease had elevated oxidised RNA but these neurons were not affected in patients with Alzheimer's disease. Further, oxidation of RNA does not interfere with the ability of the mRNA to associate with polysomes but it does cause translation errors such that short polypeptides are produced (Tanaka *et al.*, 2007). Analysis of oxidised mRNA in primary neuronal cultures revealed that oxidised mRNAs had a corresponding lower level of protein (Shan *et al.*, 2007). Indeed analysis of oxidised mRNA species in a transgenic mouse model of ALS that have a mutation in superoxide dismutase 1 (SOD1) revealed that a specific set of transcripts were oxidised (Chang *et al.*, 2008). An antibody was used to separate oxidised mRNAs from non-oxidised mRNAs isolated from spinal chord tissue, which were then analysed by microarray hybridisation. This identified a subset of transcripts, some of which have been associated with ALS disease progression, including SOD1, dynactin 1 and vesicle-associated membrane protein 1.

To determine whether RNA oxidation was responsible for the neurodegeneration and associated changes in transcript levels observed in *Adar*^{5G1} flies I used an antibody directed to 8-OHG to separate oxidised RNA from non-oxidised RNA. This antibody was raised to the product that is generated following oxidation of guanosine (8-oxoguanosine, 8-OG), however the antibody cross reacts with 8-oxodeoxyguanosine in DNA; therefore the immunoprecipitation was performed on purified DNase I treated RNA. Following the immunoprecipitation cDNA was

generated with oligo (dT) primers to select for mRNAs, and three dilutions were used in the subsequent PCR reactions to determine the presence of specific transcripts in the oxidised and non-oxidised fractions without saturating the PCR reaction.

Figure 4.15 shows the results obtained for *Ilp3* and *GAPDH* analysed with this method. These transcripts were analysed in RNA extracted from w^{1118}, $Adar^{5G1}$ and $Adar^{5G1}$*:Cha-GAL4>UAS-dAdar* flies. Quantification of the gel images was performed using ImageQuant software (GE Healthcare). Band intensity was normalised to input for each sample to account for overall differences in expression. The ratio of oxidised RNA: non-oxidised RNA was then calculated for each transcript in each fly strain, and the results are presented in a graph (Figure 4.15). Surprisingly these calculations revealed that the fraction of oxidised RNA in $Adar^{5G1}$ flies was lower than that seen in w^{1118} flies and flies expressing the *dAdar* transgene for both transcripts analysed, indicating RNA oxidation does not play a role in the neurodegenerative phenotype observed in $Adar^{5G1}$ flies. However, as RNA oxidation has been shown to be specific for the neuronal populations that undergo degeneration, a more detailed analysis of RNA isolated from neurons in the mushroom bodies may be more informative, although technically challenging.

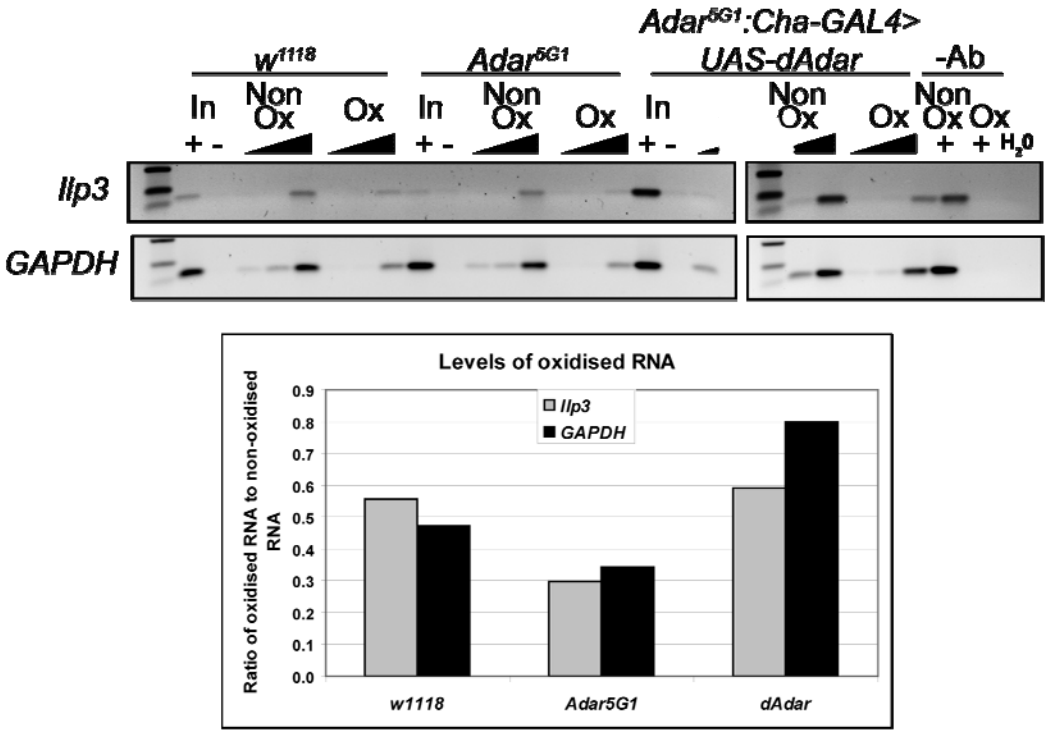


Figure 4.15: Analysing levels of RNA oxidation in *Adar*-null flies

A) To determine whether levels of oxidised RNA were altered in *Adar*-null flies an RNA immunoprecipitation was performed on RNA isolated form 5 day old fly heads with an antibody which detects 8-oxoguanosine. Oxidised RNA was bound by the antibody whilst non-oxidised RNA remained in solution. RT-PCR was performed on these two RNA species and *GAPDH* and *Ilp3* were analysed by PCR. Dilutions of cDNA were used (1/50, 1/25, and neat) to ensure the PCR reaction was in the linear phase. As a control w^{1118} RNA was incubated without antibody, and signal was only detected in the non-oxidised fraction indicating no RNA was precipitating on the beads. B) Quantification of the bands using ImageQuant software revealed a decrease in the ratio of oxidised RNA: non-oxidised RNA for both transcripts in $Adar^{5G1}$ flies.

4.3 Discussion

Recently *Drosophila* with mutations in the *blue cheese* gene, which develops progressive neurodegeneration (Finley et al., 2003), was shown to have impaired lysosomal trafficking (Lim and Kraut, 2009). A genetic screen for modifiers of the *blue cheese* neurodegenerative phenotype identified components of the autophagic pathway including *Atg1* and *Atg8a* indicating these genes are active in a common autophagy pathway (Simonsen et al., 2007). However there are differences between these mutants and *Adar*-null flies, notably *Drosophila* that have mutations in autophagy genes show accumulation of ubiquitin positive aggregates in neurons, which increases with age. This has not been addressed definitively in *Adar*-null flies; some electron dense inclusions were observed in electron micrographs of *Adar*-null brain but these require further investigation (Leeanne McGurk, Thesis). Indeed, overexpression of a key gene in the autophagic pathway *Atg8a* in wild type flies can extend the normal lifespan and reduce the occurrence of age-related neuronal inclusions (Simonsen *et al.*, 2008), indicating the processes of autophagy and neurodegeneration are linked.

Interestingly flies overexpressing *Atg8a* also show increased resistance to oxidative stress, which is also observed in *Adar*-null flies (Chen et al., 2004). This shows that upregulation of the autophagic pathway can increase resistance to oxidative stress, which could be occurring in *Adar*-null flies. However, known regulators of autophagy were not returned in pattern searches of the microarray data, and GO terms associated with autophagy were not found to be overrepresented. As these search processes are dependent on informative genome annotation, and a large number of transcripts returned in pattern searches were unannotated, it cannot be excluded that transcripts involved in the autophagy pathway are regulated by ADAR. Indeed, this would fit with the observation that ADAR regulates transcripts involved in stress response. Further work analysing the role of ADAR in response to different environmental stressors such as starvation may help to elucidate this. In addition, comprehensive analysis of gene networks which are altered may provide a better understanding of the role of ADAR in the *Drosophila* brain.

The four phenotypes associated with *Adar* deletion in *Drosophila* which are investigated herein are reduced viability, decreased and uncoordinated locomotion, loss of editing, and age-dependent neurodegeneration. Using the GAL4 binary expression system (Brand and Perrimon, 1993) to express an *Adar* transgene in the cholinergic neurons restores all four phenotypic defects to levels which are comparable to wild type strain w^{1118}. However expression of a catalytically inactive transgene does not restore editing, viability or locomotion defects, but suppresses the neurodegenerative defect. This allows dissection of the phenotypic components into editing-dependent phenotypes, which include reduced viability and the locomotion phenotype, and the editing-independent phenotype of age-dependent neurodegeneration. Many of the edited substrates of *Drosophila* ADAR are involved in synaptic transmission, indicating site-specific editing may account for the locomotion defects observed in *Adar*-null flies. To further characterise the editing-independent effects of ADAR in suppression of neurodegeneration microarray analysis was performed.

According to the microarray data ADAR expression influences the expression of several cytochrome genes which are known to be regulated in response to stress. This implicates ADAR as a regulator of stress response, analogous to the role ADAR2 has in the brain. In mouse CA1 pyramidal neurons which are vulnerable to degeneration following ischemic insult, ADAR2 has a role in mediating cell death through decreased editing of the GluR-B Q/R site which alters the permeability of the channel to calcium (Peng et al., 2006). ADAR2 was shown to be regulated by CREB transcription factor, indicating it is an effector of cell death rather than a regulator. *Adar*-null flies show increased sensitivity to hypoxic stress which mirrors the hypoxic environment in mouse CA1 pyramidal neurons following ischemic insult. What remains unclear is why *Adar*-null flies show increased resistance to oxidative stress induce by paraquat treatment (Chen et al., 2004), indicating responses to hypoxic stress and oxidative stress are mediated through different pathways.

Although extensive details were not provided, Wang *et al*, (2004) indicated that microarray analysis of RNA extracted from E11.0 ADAR1-null mice revealed anti-apoptotic transcripts *Bcl2-like 10* and *Bcl-x* were 13-fold and 4-fold decreased, respectively (Wang et al., 2004). In the microarray presented here there was no significantly over-represented GO terms associated with apoptotic cell death, and known regulators of the process were not found to be altered in the pattern analysis. This supports previous TUNEL staining data showing that apoptotic cell death is not the mechanism of neuronal degeneration seen in $Adar^{5G1}$ flies (Leeanne McGurk, Thesis).

Discovering an alternative function of ADAR may reveal an as yet undetermined role for ADAR during development. Expression analysis revealed the presence of *Adar* transcripts during early embryogenesis which are lost during the mid blastula transition. This expression pattern is frequently observed with maternally inherited transcripts, indicating ADAR could have a function at a very early stage of development (Chen *et al.*, 2009). However, initial investigation of *Adar* deletion mutants did not detect any phenotypic differences in the progeny from homozygous or heterozygous females indicating this function is not essential (Palladino et al., 2000b). Indeed multiple isoforms of *Adar* are expressed in the developing oocyte (Marcucci et al., 2009). The viability of *Adar*-null flies is approximately 20%, it is unclear at what stage of development this defect arises, however it is not restored by expression of an inactive *Adar EA* transgene indicating the viability defect is associated with editing. Further work to determine the role of ADAR during development is required.

Subsequent analysis of ADAR antagonism of the RNAi pathway in the *Drosophila* eye revealed overexpression or deletion of *Drosophila Adar* did not perturb RNAi (Heale et al., 2009). This study utilised expression of a cytoplasmic *white* hairpin RNA which directs silencing of the *white* gene producing a red eye phenotype. Expression of human ADAR1 p150 which is a cytoplasmic ADAR antagonised RNAi to produce a red eye phenotype; however human ADAR1 p110 or ADAR2 did not affect RNAi likely due to their nuclear localisation. Partial antagonism of RNAi

was observed in flies expressing a catalytically inactive ADAR p150 indicating ADAR binding alone can affect siRNA production from dsRNA. Interestingly the *white* hairpin RNA was found to be endogenously edited by *Drosophila* ADAR to a low level (<10%) indicating ADAR does interact with the *white* RNA hairpin in the nucleus prior to cytoplasmic export and processing by DICER (Heale et al., 2009). A more comprehensive analysis using flies with mutations in several loci required for small RNA production crossed to *Adar*-null flies confirmed that the locomotion defects associated with loss of *Adar* do not arise due to aberrant RNAi (Jepson and Reenan, 2009).

Immunoprecipitation of FLAG-tagged ADAR from *Drosophila* S2 cells was performed as it provided a method for screening novel candidate transcripts identified in the microarray analysis for interaction with ADAR. Constructs containing wild type *Adar* were created by a former lab member and conditions for induction were determined (Angela Gallo, unpublished data) for experiments to analyse editing levels in known substrates of ADAR. However using this system for RNA-immunoprecipitation proved technically difficult. There are several factors which affected this experiment, most notably the low transfection efficiency of the *Drosophila* S2 cells in conjunction with no way to isolate cells expressing ADAR made these experiments very difficult. In the future a cell line which endogenously expresses epitope-tagged ADAR would be beneficial. Stable cell lines generated from *Drosophila* S2 cells are able to express ectopically encoded proteins for a long period of time (Towers and Sattelle, 2002). Alternatively, transiently-transfected cells which express a GFP-tagged ADAR protein to allow fluorescence activated cell sorting (FACS) analysis prior to immunoprecipitation could be used.

A neuronal cell line may be more representative of endogenous ADAR expression; several have been described including the Kc cell line which is thought to be neural or glial in origin (Echalier and Ohanessian, 1969). Strains derived from the Kc line such as Kc-167 and Kc-H can be further differentiated into cells of a more neuronal appearance which express acetylcholinesterase by the addition of ecdysone (Cherbas *et al.*, 1977). One possibility discussed was performing cross-linked

immunoprecipitation (CLIP) (Ule *et al.*, 2003), which has been optimised for low amounts of precipitated RNA (Wang *et al.*, 2009). However this method requires extensive optimisation. Ideally this would be performed *in vivo* but endogenous *Adar* is expressed to a very low level and obtaining sufficient material may prove difficult. However, as the catalytic mutant *Adar EA* transgene is expressed to a higher level than wild type this may be of use in the identification of ADAR substrates. Preliminary work revealed the increased expression observed with the *Adar EA* transgene results in higher levels of ADAR EA protein (data not shown).

The RNA-binding mutant used in these experiments was chosen to disrupt RNA-binding without affecting the structure of the protein (Ramos et al., 2000; Valente and Nishikura, 2007). However, when the ADAR RRM-FLIS6 protein was overexpressed in *Pichia pastoris* prior to purification for biochemical analysis the protein was expressed to approximately half the level of wild type ADAR (see Appendix I). The same result was obtained when overexpressing FLAG-tagged ADAR RRM in *Drosophila* S2 cells. This is unusual as ADAR mutants that are unable to edit substrates are usually expressed to a higher level than wild type ADAR (unpublished observations). This indicates the ADAR RRM mutant protein may not be an ideal control for these experiments, as it may not be folded correctly.

Attempts to characterise the *in vitro* binding of ADAR to several candidate transcripts that were shown to be altered according to the microarray failed to establish a difference in binding between altered transcripts and control transcripts, although some specificity was observed as ADAR does not bind to yeast tRNA. It is likely that this reflects an innate RNA-binding function of ADAR, indeed it has been observed *in vitro* for ADAR2, although this was an artificial substrate generated from a known substrate (GluR-B) and a double-stranded RNA region from the potato spindle tuber viroid (Klaue *et al.*, 2003). Interestingly, promiscuous editing was observed here, likely due to the perfect dsRNA used which is known to be hyper-edited by ADARs. This study employed scanning force microscopy to allow visualisation of ADAR bound to RNA, as previously binding and editing were analysed by the presence of inosine within the RNA following interaction. Therefore

the data presented in this chapter is the first investigation of ADARs interaction with pre-mRNAs that are not known to be substrates of ADAR. This data indicates that ADAR has a wide range of binding substrates, yet only edits a select few. This specificity may be imparted by the deaminase domain.

Taken together these results indicate that *Adar*^{5G1} flies develop age-dependent neurodegeneration with lower levels of transcripts that are altered in response to oxidative stress, and lower levels of oxidised RNAs. This could be interpreted as showing that *Drosophila Adar* has a role in response to oxidative stress similar to the role of ADAR2 in mouse CA1 pyramidal neurons. However although capable of interacting with candidate transcripts *in vitro* it is unclear whether ADAR interacts with these transcripts *in vivo*, and therefore the possibility of an effect on transcription cannot be excluded.

Chapter 5: Analysis of alternative splicing in *Adar*-null flies

5.1 Introduction

As described in Chapter 4 microarray analysis performed on RNA extracted from *Adar*-null flies revealed a class of transcripts which were miss-regulated in the absence of ADAR. However it remains unclear how ADAR is affecting the level of these transcripts and how this contributes to the neurodegenerative phenotype.

ADAR functions as a RNA-editing enzyme, and flies which lack editing due to deletion of *Adar* show locomotion defects and develop age-dependent neurodegeneration (Palladino et al., 2000b). However, the age-dependent neurodegeneration phenotype can be suppressed by expression of a catalytically inactive *Adar EA* transgene in the cholinergic neurons, indicating ADAR has a deaminase-independent function. ADAR contains a conserved deaminase domain which is mutated in the catalytically inactive *Adar EA* transgene, while the two dsRBDs remain intact in this mutant. Therefore, to characterise the editing independent function of ADAR I focused on biological processes which require RNA-binding.

Alternative splicing (AS) is the process by which several different transcripts can be produced form the same pre-mRNA due to inclusion or exclusion of exons (Black, 2003). AS is a mechanism of generating protein diversity without expanding the genome, and it can change the function of the protein produced or alter the regulation of expression through use of different promoter or UTR regions. There are several ways in which RNA editing could impact on AS. Firstly, editing by ADAR could alter regulatory sequences required for AS including the consensus 5' and 3' splice site sequences which both include adenosine residues. Therefore A-to-I editing could abolish existing splice sites or create new splice sites which is demonstrated in the rat ADAR2 autoediting event (Rueter et al., 1999), and the *PTPN6* transcript (Beghini et al., 2000). Editing could also strengthen or weaken splicing enhancer or silencer sequences which can be located in either exonic or intronic regions. However as many splicing enhancer or silencer sequences are poorly defined changes induced through editing these sequences may be subtle, and alterations in these elements through A-to-I editing has not been well characterised to date.

ADAR binding could also influence AS through blocking access of splicing factors to splice sites. Inhibiting the access of splicing factors in this way would only require RNA-binding and is therefore independent of editing events. Interestingly the catalytically inactive ADAR which is able to rescue the neurodegenerative phenotype associated with deletion of *Adar* retains RNA-binding activity. Editing independent inhibition of miRNA processing has recently been shown to occur as ADAR binding can inhibit cleavage of miRNA hairpin precursors by DROSHA or DICER ribonucleases (Heale et al., 2009; Kawahara et al., 2007a; Yang et al., 2006). The dsRNA structure required for editing often forms between exonic and intronic sequences indicating editing must occur before splicing, therefore binding of ADAR to pre-mRNA may alter splice site use or processing of introns.

Finally, A-to-I editing could alter the secondary structure of an RNA molecule, either through stabilising a dsRNA stem-loop structure or weakening it. The ECS required for editing by ADARs has only been determined for some *Drosophila* ADAR substrates and these structures can contain looped regions of several kilobases or be complex pseudoknots. Investigation of the *para* transcript isolated from flies which lack the RNA helicase encoded by *maleless* revealed aberrant editing and splicing catastrophe with large sections of the transcript which show extensive secondary structure absent from the mature message (Reenan et al., 2000). This indicates that the secondary structure required for ADAR editing may be detrimental to the splicing process in some transcripts.

Several studies have been performed to analyse the association between RNA editing and alternative splicing, yet no consensus has been reached (Agrawal and Stormo, 2005; Bratt and Ohman, 2003; Flomen et al., 2004; Jin et al., 2007; Jones et al., 2009; Schoft et al., 2007). Specifically the hypothesis I wanted to investigate was whether aberrant alternative splicing could account for the difference in transcript levels observed with the gene expression microarray analysis described in Chapter 4.

There have been extensive attempts to characterise splicing isoforms at specific developmental time points, in different tissues, or in response to different external

stimuli. A recent technological advance which has aided these studies is the use of microarrays designed to recognise specific splice forms. Global analysis of AS events in tissues or whole organisms using microarrays has allowed specific questions about alternative splicing and its regulation. There are several different types of AS-sensitive microarrays that are routinely used, which are reviewed in (Moore and Silver, 2008). Briefly, exon arrays detect all known exons but cannot distinguish specific isoforms therefore require extensive validation to determine specific isoform differences. Tiling arrays utilise probes spanning the entire genome or the genomic region around transcripts of interest and are capable of detecting novel exons and retained introns which other arrays may miss. However tiling arrays generate large datasets making analysis difficult and interpretation of results can be confused by non-coding transcripts. Splice-junction arrays contain many probe sequences per transcript, some specific to exon-exon junctions whilst others lie within constitutive exons which detect background expression levels to allow correction for overall differences in transcript level. These arrays require prior knowledge of AS and therefore cannot identify novel splicing events.

AS arrays have been used successfully to identify targets of the mammalian neuronal splicing protein NOVA (Ule *et al.*, 2005) indicating this technique can be used to analyse complex splicing patterns found in the brain. Splice junction arrays have also been used to analyse the effect of siRNA knockdown of the homologs of mammalian hnRNP A/B family in *Drosophila* S2 cells (Blanchette *et al.*, 2009). hnRNP proteins are specific regulators of AS and depletion of individual members of the family revealed combinatorial regulation with multiple hnRNP proteins binding to the same transcript. Investigation of responses to external stimuli was also investigated using splice-junction arrays. AS changes associated with activation of the wingless signalling cascade were analysed alongside changes induced by insulin treatment (Hartmann et al., 2009). These pathways have been extensively studied for effects on transcript levels, however little is known about how AS events are regulated in response to signalling cascades. Analysis of these pathways in parallel revealed distinct transcripts were subject to regulation at the levels of splicing (Hartmann et al., 2009).

A custom designed AS array was used to analyse global changes in splicing in RNA extracted from flies lacking the Adar gene. This identified a specific set of genes which are aberrantly spliced in Adar-null flies. These genes were distinct from the genes which showed altered transcript levels according to the Affymetrix array and few known ADAR substrates were included in this list. Real time PCR with primers which spanned the affected junctions confirmed the AS array data. 3' Rapid Amplification of cDNA Ends (RACE) was used to analyse the transcripts from these genes in w^{1118}, $Adar^{5G1}$ and $Adar^{5G1}$*:Cha-GAL4>UAS-dAdar* flies and this revealed that transcripts isolated from $Adar^{5G1}$ flies used alternative polyadenylation sites. These alternatively polyadenylated transcripts were not observed in RNA extracted from wild type flies or in flies expressing an *Adar* transgene in the cholinergic neurons. Investigation of the AS array data indicated this phenomena may be more widespread although further work is required to characterise the use of alternative polyadenylation in *Adar*-null flies.

5.2 Results

AS array design

To determine the profile of AS changes in *Adar*-null flies global analysis of AS was performed using a custom-made *Drosophila melanogaster* splice junction array in the laboratory of Marco Blanchette, Stowers Institute for Medical Research, Missouri, USA. The AS array and the analysis package named Sparrow were designed and run by Marco Blanchette, and analysis was performed using Sparrow run in the R software environment.

The array was designed to analyse splice junctions in 13,344 genes, however not all of the genes included are known to be alternatively spliced. The array comprises three overlapping probes which span each exon-exon junction, and probes within constitutively spliced exons to determine expression levels (Figure 5.1). However when analysis was performed the differences in expression levels were normalised so

only differences in the exon-exon junctions used were calculated. This allows a comprehensive analysis of alternative splicing differences, but also provides information on expression of transcripts. Three tiled probes per junction were used to maximise the chance of obtaining data for each junction. In contrast to expression or exonic arrays, junction array probes cannot be chosen for optimal probe design as they must span the junction to be analysed. In the past junction arrays have utilised one probe per junction and this resulted in large amounts of missing data. For statistical analysis all three probes must be detected, however if fewer than three probes are detected the remaining data is still plotted as this can be informative although not significant.

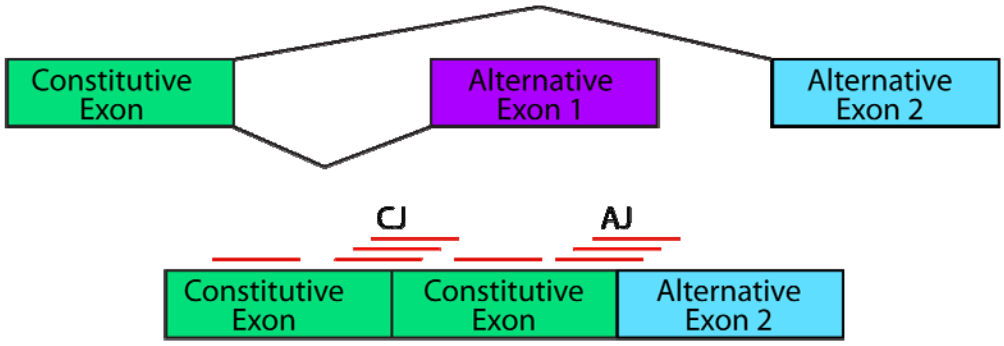


Figure 5.1: AS microarray design
Probes were designed to span both constitutive junctions (CJ) and alterntive junctions (AJ) along with internal exonic probes within constitutive exons to analyse background expression levels. Exon junctions are tiled with three probes which are offset by 3nt to try and ensure a read is obtained from each junction.

RNA preparation for AS Array

RNA was isolated from 100 heads of 5 day old flies from the w^{1118} and $Adar^{5G1}$ strains to allow direct comparison with the Affymetrix expression microarray data. Due to the cost of the arrays the AS array analysis was only performed on w^{1118} and $Adar^{5G1}$ RNA. RNA extraction was performed with TRIZOL reagent and protocol (Invitrogen), DNase I treated (Ambion) and purified using a Qiagen Minelute column (Qiagen). RNA samples were sent to Marco Blanchette at the Stowers Institute for Medical Research in the USA for cDNA amplification and hybridisation

to the microarrays. Prior to microarray analysis RNA was analysed on an Agilent Bioanalyzer (Agilent Technologies) to determine the quality and quantity. The results of the Agilent Bioanalyzer are shown in Figure 5.2. This indicated that one of the *Adar*^{5G1} biological replicates was significantly degraded (Adar_5G1_B). During purification the pellet in sample Adar_5G1_B appeared discoloured, indicating there was a problem and this sample may have been contaminated. A fresh batch of 100 *Adar*^{5G1} fly heads were collected and used for analysis.

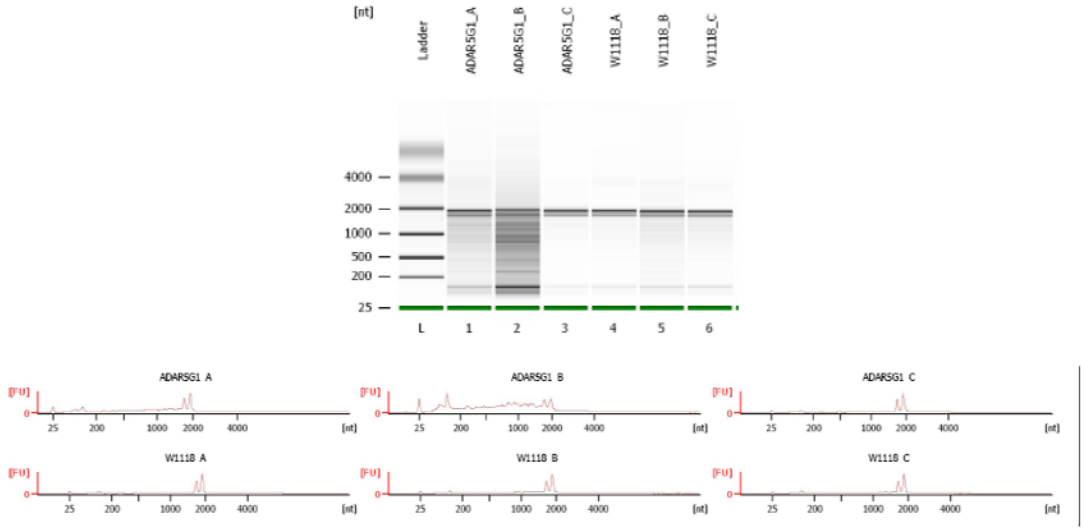


Figure 5.2: Agilent Bioanalyzer results of RNA integrity
Total RNA was purified from *w*^{1118} and *Adar*^{5G1} fly heads using TRIZOL reagent and protocol (Invitrogen) followed by DNAse I treatment (Ambion) and concentration using a Qiagen MinElute column (Qiagen). Integrity of the purified RNA was analysed on an Agilent Bioanalyzer (Agilent Technologies) before it was amplified for hybridisation to the custom-made AS array. One RNA sample from *Adar*^{5G1} flies was degraded (Adar5G1_B) possibly due to the time taken to collect the fly heads, and therefore was not used. A replacement RNA sample was used in further analysis (data not shown). The top panel shows the gel image, the lower panels show electropherograms from each lane of the gel.

AS array results

For hybridisation the *w*^{1118} cDNA was pooled to create one control sample and each of the three biological replicates of *Adar*^{5G1} cDNA were compared to this control to minimise variation. *w*^{1118} and *Adar*^{5G1} cDNA samples were labelled with different fluorescent dyes and both were hybridised to the arrays at the same time. This is a different procedure to the Affymetrix arrays where only one sample is hybridised at a

time and the relative quantities are calculated *in silico*. The arrays were scaled and normalised using the Lowess method in the marray package in R. Genes were analysed individually with an ANOVA test to determine whether any of the junction probes for a given gene were significantly altered from the exonic probes for that gene. This generates a list of genes which contain significantly altered junctions. A t-test was then performed to find the significantly altered junctions within these genes and multiple testing corrections were applied. Finally, to determine the expression of the exon junctions, a net expression (NE) value was calculated from the common exonic probes which are constitutively present, and this was subtracted from the junction probe expression data. This highlights differences in exon junction use irrespective of changes in the NE value between samples. This analysis was run using commands built in to the Sparrow package.

Analysis indicated that signal was detected for 12,311 genes out of 13,344 analysed. Of these, 759 genes analysed showed significantly altered splicing using the cut off of a p value of 0.001. A comprehensive list of these genes can be found in Appendix II. A summary list of the 10 most significantly altered genes is shown in Table 5.1. The table shows that 4 of the 10 genes listed are not annotated as being alternatively spliced. This could indicate novel isoforms of these genes have been detected or cryptic splice sites utilised. Of the six genes which are known to be alternatively spliced only three showed altered junction usage at junctions involved in alternative splicing events. The remaining seven genes showed changed levels at constitutively spliced junctions. This was not expected as the majority of the genes included in the array are alternatively spliced. Taken together the analysis of the most significantly affected genes indicate that ADAR is not a major component in splice-site choice during alternative splicing, however it does suggest a role for ADAR in constitutive splicing events.

Symbol	Gene Name	Function	Splicing status
dp	Dumpy	electron carrier; endonuclease	not known to be AS
Strn-Mlck	Stretchin-MLCK	myosin light chain kinase	different isoforms affected but not AS jxns
Syd	Sunday driver	kinesin binding	alternative first exon affected
ord	orientation disrupted	protein binding	not known to be AS; many probes failed
CG13900		damaged DNA binding	different isoforms affected but not AS jxns
Adh	Alcohol dehydrogenase	alcohol metabolism	locus overlaps with Adhr
TepIII		phagocytosis (*S.aureus*)	not known to be AS
CG31190	Dscam3?	cell adhesion	different isoforms but not AS jxns
Sap47	synapse assoc 47kDa	synaptic transmission	alternate 2nd exon affected; but all possibilities over-represented
CG3734		serine-type peptidase	not known to be AS

Table 5.1: Details of the 10 most significantly altered genes from the AS array analysis of RNA from *Adar*-null flies.
Four of the 10 most significantly altered genes are not known to be alternatively spliced (AS). Of the remaining six which are known to be AS only three are affected at junctions which are alternatively spliced, the remaining 7 genes have altered splicing at constitutive junctions.

The AS array data can be visualised in several ways. Initially graphs were generated plotting the log ratio of $Adar^{5G1}$ to w^{1118} for each junction across a gene. Figure 5.3 shows the graphs of the 12 genes which showed the most significant alteration in alternative splicing. The junction probes are listed along the x-axis and the common (constitutive) exonic probes are listed at the left hand side of the graph for comparison. Red coloured probes indicate highly significantly altered probes ($p < 0.001$), and pink are mid-significance probes ($p = 0.01$-0.001). The y-axis represents fold change in expression between $Adar^{5G1}$ and w^{1118} samples.

A striking observation from the graphical representation of the data is the variation across the common exonic probes. The common exonic probes are placed in constitutive exons throughout the gene and therefore there may be differences in signal intensity between 5' and 3' region due to differences in amplification. The cDNA was generated using oligo (dT) primers and the 3' region of transcripts may have been amplified preferentially using this method. However, oligo (dT) primers were used to specifically amplify polyadenylated mature RNA transcripts from total RNA, and any variation across transcripts should be similar in the RNA samples analysed. The variation appears higher across transcripts with low signals which is consistent with previous AS array data (Marco Blanchette, unpublished observation).

Three probes were analysed across each junction to try and optimise the likelihood of obtaining a consistent signal as probe design is limited to the region spanning the exon-exon junction and therefore may not be optimal. However for some junctions analysed one or more probes failed leaving gaps in the data. The probes that gave signals for these junctions are included as they may be informative, but the results are not statistically significant.

Another striking observation was that many transcripts showed completely separate junctions being altered, often with many unaltered junctions in between. This was an unexpected result as inclusion or exclusion of exons should produce an alteration in two or more junctions surrounding the exon which is affected. However this could indicate that cryptic splice sites are used which are not annotated on the fly genome and therefore were not included on the array. Another explanation is that the processing of introns is altered or inefficient and this is observed by microarray analysis as a decrease in junction use. Therefore further investigations focused on analysing altered transcripts to determine whether they contained cryptic splice sites or retained introns.

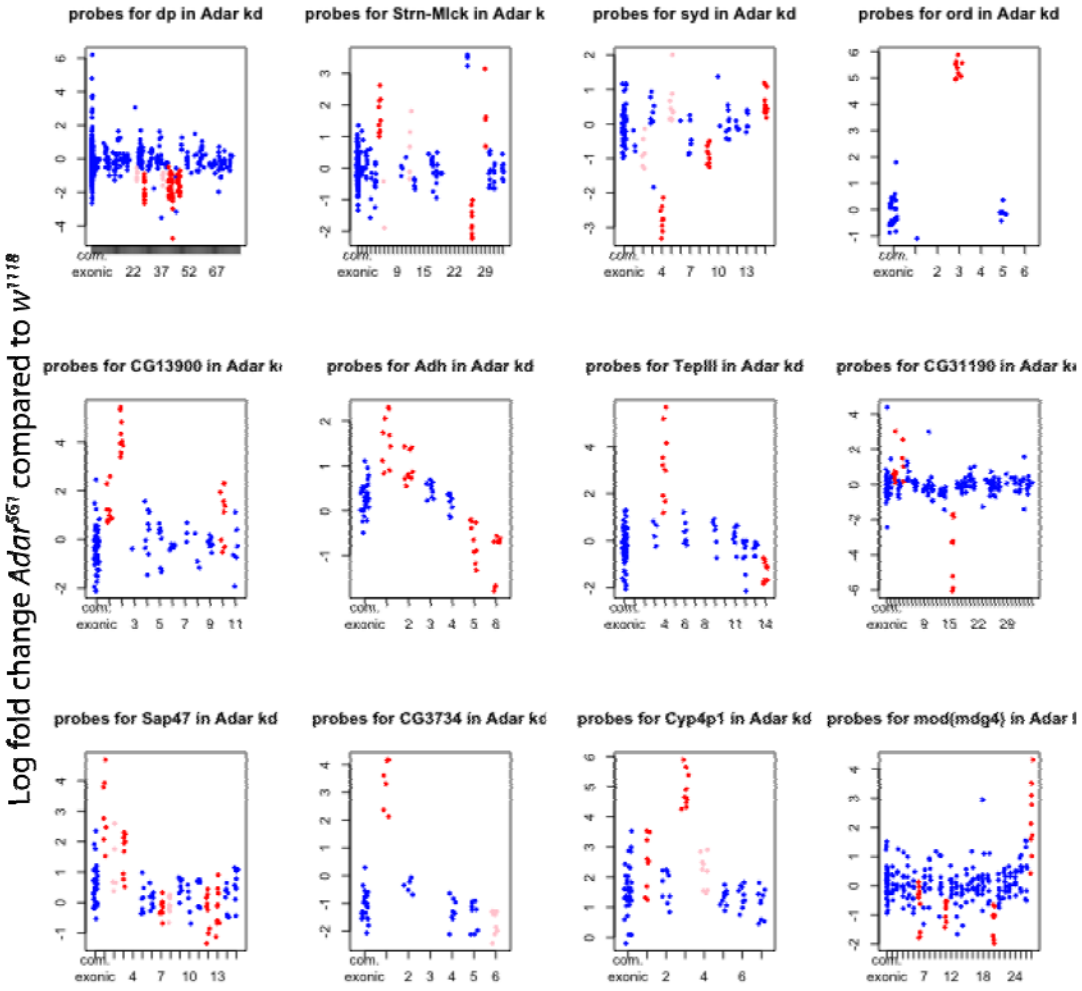


Figure 5.3: Graphical representation of junction probes across the 12 most significantly altered transcripts.
The x-axis displays the junctions across the transcript which are numbered 5' to 3' and comm exonic shows the common exonic probes used to determine the expression. The y-axis shows the log fold change of *Adar*^{6G1} compared to *w*^{1118}. There are three probes per junction and three biological replicates therefore there are nine data points per junction. For statistical analysis junctions where probes failed are not included, therefore some junctions appear to be significantly different from other junctions within the gene but are coloured blue because they did not meet the criteria for statistical analysis. Where red indicates a the p value of the t-test is less than 0.001 (highly significant), pink indicates the p value is between 0.01 and 0.001 (mid significance), and blue indicates all non-significant junction probes and common exonic probes.

Another script developed by Marco Blanchette during the AS array analysis allowed the AS array data to be superimposed onto the UCSC genome browser (http://genome.ucsc.edu). This was very useful for visualising the affected junctions within a gene and determining which isoforms were affected. Previously only the coordinates of the affected junctions were available and primer design was difficult, therefore visualising the affected region on the genome browser was very useful. Examples of two affected genes visualised on the UCSC genome browser are shown in Figure 5.4. Two different tracks could be manually added to the genome browser, one mapping the junction probes used with their corresponding numbers, and one mapping the fold change difference in signal between wild type and *Adar*-null flies, although significance is not shown on this display. The scale used for the fold change data varies for each transcript. The UCSC genome browser also displays cloned expressed sequence tags for each locus allowing a clear view of the affected isoforms.

The examples shown are *CG13900* which encodes SF3b120 a component of the U2 snRNP (Mount and Salz, 2000) and *CG31190* which is also known as *Dscam3* (Andrews *et al.*, 2008). *CG13900* contains one splice junction which is highly overrepresented in *Adar*-null flies (4.32 fold), despite the neighbouring junctions being present at similar levels in both strains. The affected junction is a constitutive junction which is located within the 3'UTR region and is present in both transcripts annotated from this locus. Analysis of the expression data from the constitutive exonic probes within *CG13900* revealed there was no change in expression of these isoforms in *Adar*-null flies. This example demonstrates how some of the data obtained did not fit the expected pattern of a change in known isoform levels and may indicate another isoform is expressed from this locus which is not annotated on the *Drosophila* genome and is therefore not analysed by the AS array. Alternatively, the probes which analyse the overrepresented junction could cross-hybridise to another transcript although this would have to be differentially expressed between wild type and *Adar*-null flies to produce a signal like the one shown. A BLAST search of the region targeted by the affected exon-junction probes did not reveal any significant homology to transcribed *Drosophila* sequences in the Ensembl database.

It is intriguing to discover altered splicing in a gene which encodes a component of the splicing machinery in *Adar*-null flies. Many proteins involved in the RNA processing pathway are regulated at the level of alternative splicing, and this may contribute to the altered splicing seen in other transcripts. However as the expression of the *CG13900* isoforms is not altered this seems unlikely. Furthermore the U2 snRNP is a core component of the splicing machinery and we might expect to see many more transcripts affected if it were incorrectly processed. Interestingly *CG13900* contains an Interpro domain which is also present in cleavage and polyadenylation specificity factor (CPSF).

CG31190/*Dscam3* is one of four *Dscam* genes in *Drosophila* involved in cell adhesion and axon guidance (Andrews *et al.*, 2008). *Dscam3* does not show the same transcript diversity as *Dscam* (described in the Introduction section on Alternative Splicing), with only four transcripts annotated from the *CG31190* locus. One splice junction was significantly underrepresented in *Adar*-null flies (-3.06 fold). This junction is only present in one isoform of *Dscam3* indicating there may be downregulation of this isoform in *Adar*-null flies however other junctions which are shared between all isoforms are not altered. Further work is required to determine whether one *Dscam3* isoform is downregulated in *Adar*-null flies.

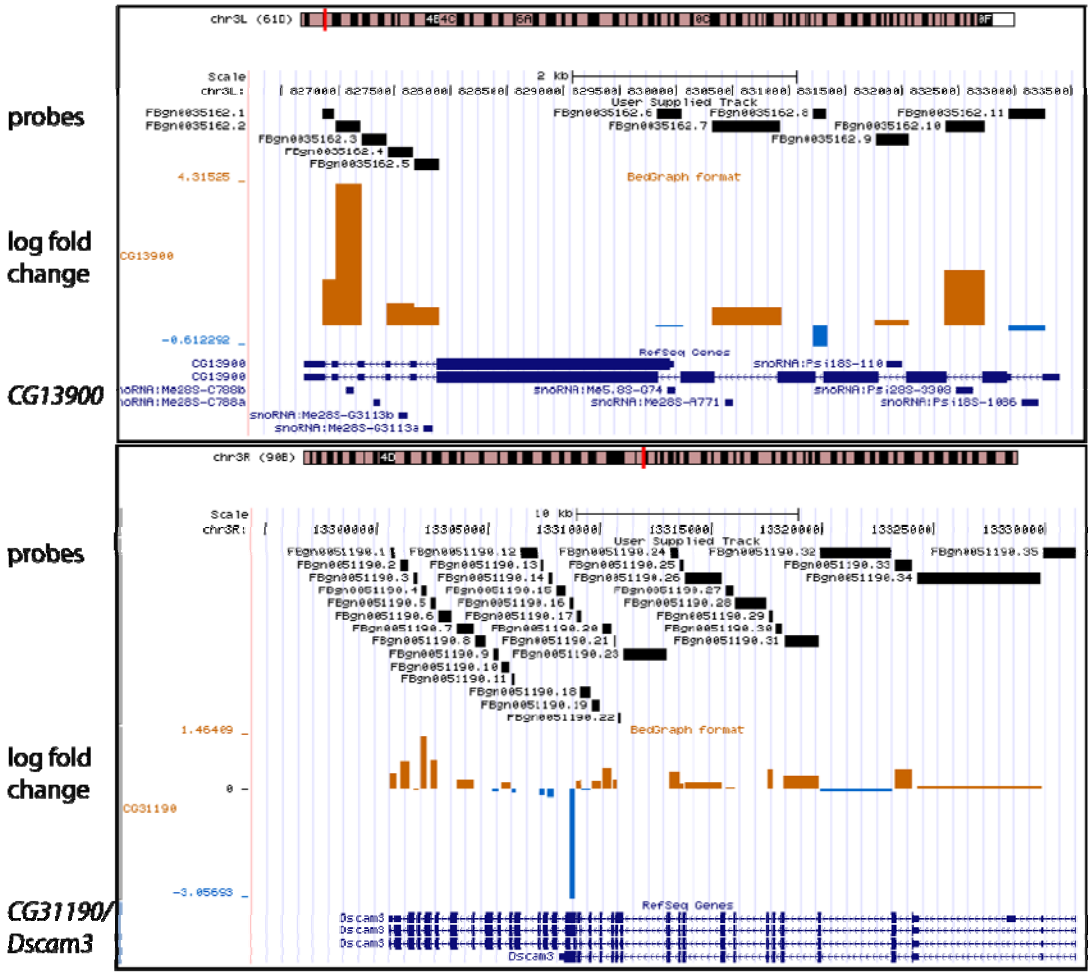


Figure 5.4: AS array data displayed on the UCSC genome browser
The AS array data can be imported into the UCSC genome browser as two separate custom tracks. One track shows the probes used in the AS array which span exon-exon junctions and their number which corresponds to the junction numbers displayed on the graphs in Figure 5.3. The other track shows the log fold change of *Adar*^{5G1} compared to *w*^{1118} samples. However, significance is not shown in this data display. The top panel shows the gene *CG13900* and the lower panel shows *CG31190* which is also known as *Dscam3*. Images are taken from the UCSC genome browser (http://genome.ucsc.edu). Both genes show one significantly altered junction. In *CG13900* a constitutively spliced junction is overrepresented in *Adar*-null flies however surrounding junctions are not significantly altered. Whereas in *Dscam3* one junction which is only found in one of the four annotated transcripts from this locus is significantly underrepresented in *Adar*-null flies indicating this isoform may be downregulated in *Adar*-null flies.

Analysis of GO terms which were significantly overrepresented in aberrantly spliced transcripts when compared to the list of genes analysed on the AS array revealed that genes involved in neurotransmission, cytoskeletal organisation and glutamate receptor activity were overrepresented. The list of overepresented GO terms is shown in Table 5.2.

The observation that genes annotated with the GO term of glutamate receptor activity are overrepresented in the AS array analysis is intriguing as glutamate receptor transcripts are known substrates of mammalian ADARs. The mammalian GluR-B transcript is edited at two sites, the Q/R site and the R/G site, which leads to changes in the amino acid sequence (Higuchi et al., 1993). Editing at the Q/R site alters the processing of the GluR-B subunits (Greger et al., 2002) and also impacts on the tetramerisation of the mature AMPA receptors (Greger et al., 2003). Q/R site editing also alters the function of the receptor by altering the calcium permeability of the multisubunit receptor and changing the rate of desensitisation (Burnashev et al., 1992). The mammalian glutamate receptor subunits are also alternatively spliced at the mutually exclusive FLIP/FLOP exons which impacts on AMPA receptor desensitisation kinetics (Koike et al., 2000; Sommer et al., 1990). Therefore it is intriguing that several transcripts encoding *Drosophila* ionotropic glutamate receptors were recovered in analysis of AS changes in *Adar*-null flies.

In total eleven genes associated with the GO term "glutamate receptor activity" were identified as aberrantly spliced in *Adar*-null flies. These include the genes *GluRIIB*, *GluRIIC*, and *GluRIIE* which encode glutamate receptor subunits that form kainate-gated receptors. Mammalian kainate-gated glutamate receptors are also ADAR substrates and both GluR-5 and GluR-6 transcripts are edited at one site which leads to an amino acid change at the position which corresponds to the Q/R site in GluR-B. GluR-6 is also edited at two other sites which lead to re-coding events. Editing levels vary throughout development and also vary in different regions of the rat brain (Bernard et al., 1999). Similar to editing of the GluR-B subunit transcript, editing of mammalian kainate-gated receptor subunits alters the permeability of the receptor to calcium. However, in *Drosophila* these transcripts are not known to be ADAR substrates, and each have only one annotated transcript indicating they are not alternatively spliced. The three aberrantly spliced genes each contain one affected junction which is annotated as a constitutive junction. However this may indicate multiple isoforms are generated by AS which are not annotated on the *Drosophila* genome. Intriguingly the *Drosophila* glutamate receptors are composed of four different subunits; they contain either a GluRIIA subunit or a GluRIIB subunit and

three common subunits (GluRIIC, GluRIID, and GluRIIE) (Qin *et al.*, 2005). Therefore the three glutamate receptor subunits which show altered splicing in the absence of ADAR form one type of kainate receptor (B-form) along with the GluRIID subunit. This suggests that ADAR has a role in the regulation of glutamate receptor subunit transcript processing, which is analogous to the role of mammalian ADARs which edit several glutamate receptor subunit transcript. However receptor composition and processing is extensively regulated and indirect effects cannot be excluded (Liebl and Featherstone, 2005).

Several other genes associated with the glutamate receptor activity GO term were also aberrantly spliced in *Adar*-null flies including *Nmdar1* and *Nmdar2* transcripts which encode NMDA-gated glutamate receptor subunits (Xia *et al.*, 2005a). These genes are alternatively spliced however they both contain only one affected junction which is a constitutively spliced junction. Interestingly the genes associated glutamate receptor activity all encode cation channels, anion channel genes were not overrepresented in this analysis.

Further work is required to determine whether ADAR binds directly to these glutamate receptor subunit transcripts and therefore regulates calcium permeable glutamate receptors in the *Drosophila* brain. It is intriguing that these transcripts show a similar profile whereby many of the glutamate receptor subunit genes (3 out of 5 mentioned here) have only one annotated transcript and all have only one affected junction which is significantly different from all other junctions. One hypothesis which could explain this pattern is that the intron at the affected junction is retained or that the junction is skipped altogether in *Adar*-null flies. Alternatively ADAR could be acting in the wild type flies to promote intron retention and in its absence the intron is efficiently spliced. Further work is required to confirm these changes detected by AS array. However these transcripts only contained single affected junctions therefore initial experiments to confirm the AS array changes focussed on genes which contained multiple affected junctions and were most significantly altered.

Genes with a tropomyosin Interpro domain were also overrepresented which could be interpreted as a role for ADAR in regulating muscle contractions however tropomyosin is also involved in cytoskeletal organisation. Genes associated with the GO term tropomyosin that were aberrantly spliced in *Adar*-null flies include *myosin heavy chain* (*Mhc*) and *Mhc1*, along with *tropomyosin* -1 (*Tm1*) and -2 (*Tm2*). These genes represent a biologically linked group of genes which are known to be alternatively spliced in a developmentally regulated and tissue specific manner, and which are aberrantly spliced in *Adar*-null flies. For example, *Mhc* contains 5 exon clusters which are spliced in a mutually exclusive manner similar to the *Dscam* locus and potentially generates 480 isoforms (reviewed in (Swank *et al.*, 2000)). In *Adar*-null flies *Mhc* shows AS changes at many exon-exon junctions, which likely represents a subtle shift in the isoforms produced from this locus. AS of *Drosophila Mhc* is developmentally and tissue-specifically regulated and altered isoform expression can lead to a flightless phenotype (Wells *et al.*, 1996). Therefore aberrant AS of genes containing tropomyosin Interpro domains may contribute to the lack of coordinated movement and flight observed in the *Adar*-null flies.

Class	GO term	Adjusted p-value
Cellular component	4: membrane-bound vesicle (GO:0031988)	0.00537018
	6: cytoplasm (GO:0005737)	0.000800717
	7: intracellular non-membrane-bound organelle (GO:0043232)	0.000250861
	7: cytoplasmic part (GO:0044444)	0.00419151
	8: cytoskeleton (GO:0005856)	0.000532567
	8: cytoplasmic vesicle (GO:0031410)	0.00196266
	8: contractile fiber (GO:0043292)	0.0129711
	9: actin cytoskeleton (GO:0015629)	0.0298812
	9: cytoplasmic membrane-bound vesicle (GO:0016023)	0.0298812
	9: contractile fiber part (GO:0044449)	0.0298812
Biological process	5: transmission of nerve impulse (GO:0019226)	0.001393
	5: cytoskeleton organization and biogenesis (GO:0007010)	0.038149
	6: synaptic transmission (GO:0007268)	0.00164881
Molecular function	4: cytoskeletal protein binding (GO:0008092)	0.000574523
	4: calmodulin binding (GO:0005516)	0.00396343
	5: actin binding (GO:0003779)	0.000138492
	5: glutamate receptor activity (GO:0008066)	0.0287268
Interpro domains	IPR000533: Tropomyosin	0.0493884

Table 5.2: Significantly overrepresented GO terms associated with aberrantly spliced transcripts.
Analysis of 759 genes which showed significantly altered alternative splicing revealed that genes involved in neurotransmission, cytoskeletal organisation and glutamate receptor activity were overrepresented.

Few edited transcripts are altered according to the AS array

As with the Affymetrix expression array analysis, few known substrates of ADAR were differently spliced in the absence of ADAR. This indicates that ADAR may be interacting with a wider range of transcripts than previously thought. Known substrates are useful in this analysis as they can be used to define the region of the transcript which ADAR binds to. Mapping editing sites onto UCSC genome browser AS array data reveals that in some transcripts the exon-junctions surrounding the edited adenosines are overrepresented in *Adar*-null flies, whereas in other transcripts the neighbouring junctions are underrepresented in flies lacking *Adar* (Figure 5.5). It is not always possible to determine the ECS region which forms the dsRNA structure required for editing as the ECS can be located up to 2 kilobases away from the edited nucleotide therefore it is unclear whether adjacent sequences are also bound by ADAR. However as with the Affymetrix array data there are only a few known substrates of ADAR which are aberrantly spliced in the absence of *Adar*. This suggests that the processes of editing and splicing are not linked for all transcripts and general conclusions cannot be drawn, rather any effects observed are specific to the transcript involved. Detailed bioinformatic analysis of the secondary structures of these transcripts is required to determine whether the affected junctions are within the dsRNA structure which is bound by ADAR.

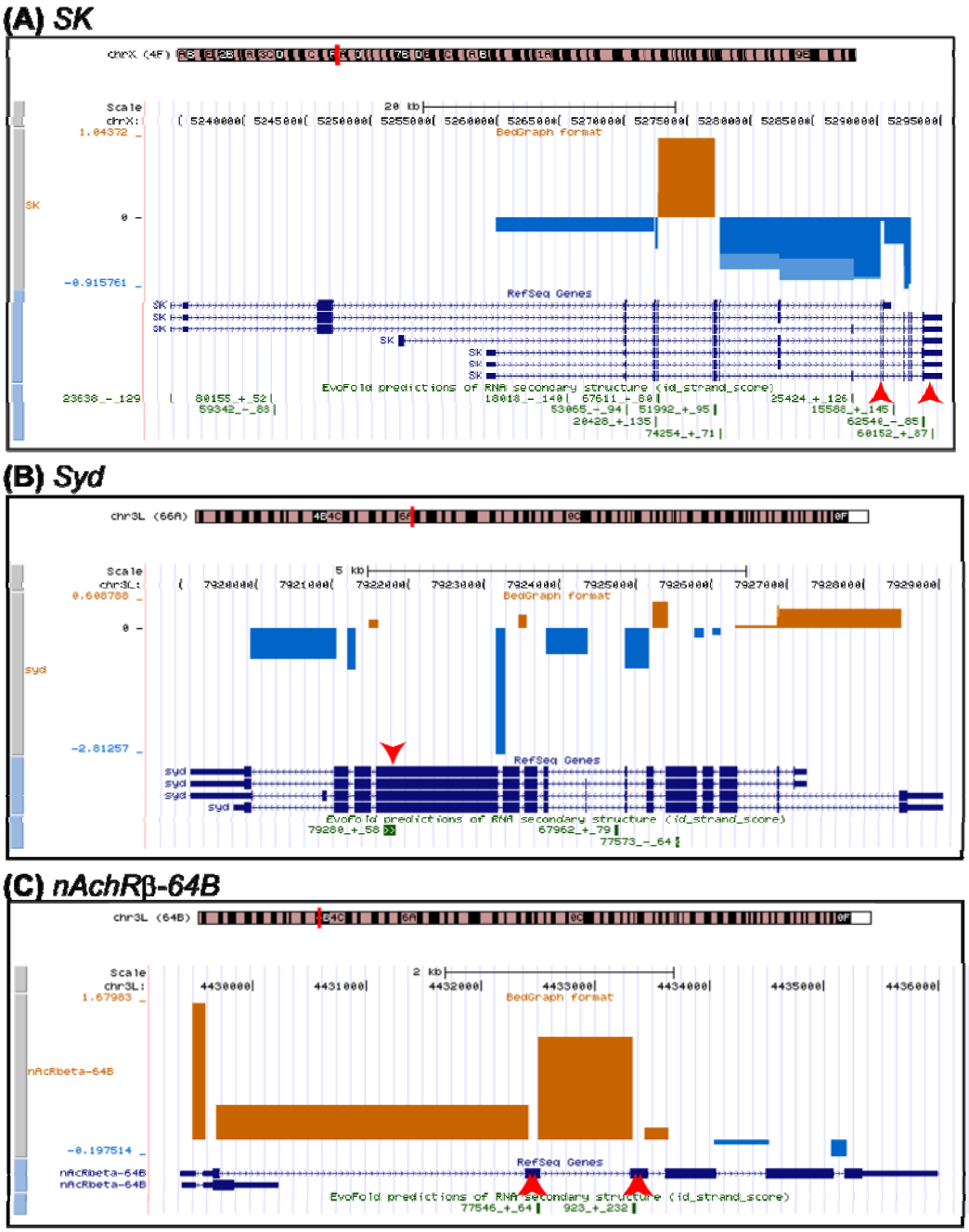


Figure 5.5: Aberrant splicing in ADAR substrates.
(A) The *SK* transcript is edited at two sites in the 3' region of the transcript (indicated by red arrowheads underneath the corresponding exons). The AS array data indicates splicing is less frequent in this region in flies that lack ADAR. (B) Similarly *sunday driver* (*syd*) is edited at one position in a large central exon, and the splice-junction at the 5' end of this exon is used significantly less frequently in *Adar*-null flies. However the splice junction at the 3' end of the edited exon is used at similar levels in wild type and *Adar*-null flies. (C) The nicotinic receptor subunit encoded by *nAchRβ-64B* is edited at four nucleotides in exons 3 and 4, however the AS array data indicates that use of splicing junctions in this region is enhanced in *Adar*-null flies. Therefore the effects of loss of editing on splicing appear to vary with different transcripts.

AS array expression data correlates with Affymetrix expression array data

Analysis of expression data obtained from the AS array allows comparisons with the Affymetrix expression array data described in Chapter 4. Analysis of the expression data obtained form the AS array revealed that 281 genes were downregulated with a $\log_{(2)}$ fold change of greater than 1.5, whilst 221 genes were upregulated with the same criteria. This is similar to the results obtained the Affymetrix arrays, however fewer genes are analysed on the AS array. Graphing the log fold change for transcripts which were called as present on both arrays showed a clear trend in the data (Figure 5.6), however the correlation was weak ($R^2 = 0.1987$). The data has a Pearson correlation of 0.45 (p-value < 2.2e-16, 95% confidence interval: 0.3750685 - 0.5112486) and the number of genes used in this analysis was 534 genes. The weak correlation was due to a large number of genes which showed upregulated levels according to the Affymetrix array but showed downregulated levels according to the expression data obtained from the AS array (highlighted with a red box in Figure 5.6).

However when the analysis was repeated limiting the genes analysed to those which showed greater than two-fold change in expression this revealed a strong correlation ($R^2 = 0.8819$), indicating that the significantly altered transcripts were consistent when analysed with both microarrays. Restricting the analysis to genes which showed over two-fold changes in transcript level gave 60 genes and a Pearson correlation score of 0.939 with a p-value < 2.2e-16 (95 percent confidence interval: 0.8997563 - 0.9633247). Therefore the transcripts with the most significantly altered expression profiles were detected by both microarrays indicating these are the transcripts which should be analysed further. The arrays differ significantly in the detection methods which could account for the observed differences in expression data. The Affymetrix array utilises 10-14 probes per gene, all located in or near to the 3'UTR of the transcript. As analysis is usually performed on oligo (dT) primed cDNA, these arrays are designed to maximise signal reads. The AS array however utilises numerous exonic probes throughout the gene to detect the basal expression

level and the data is presented as the median level for all the exonic probes. The cDNA analysed is also amplified with oligo (dT) primers which can preferentially amplify the 3' end of transcripts. However as the value used is a median transcript level any differences across a transcript due to unequal amplification of end regions would be lost. Therefore the expression data from the AS array does not show the wide range of values observed with the Affymetrix array data.

These differences can be used to explain why the AS array data does not show the same dramatic changes in expression as the Affymetrix array. However it cannot explain the group of genes which show upregulation according to the Affymetrix array but are downregulated according to the AS array. These transcripts may have different 3'UTR regions associated with them. They may be long transcripts which have an increased likelihood of differential amplification across the transcript resulting in an elevated reading for 3'UTR probes but lower expression when the whole gene is analysed. Finally, these may represent a group of genes with overlapping 3'UTR regions such that the Affymetrix probes detect multiple transcripts but the AS array probes only detect specific genes or isoforms.

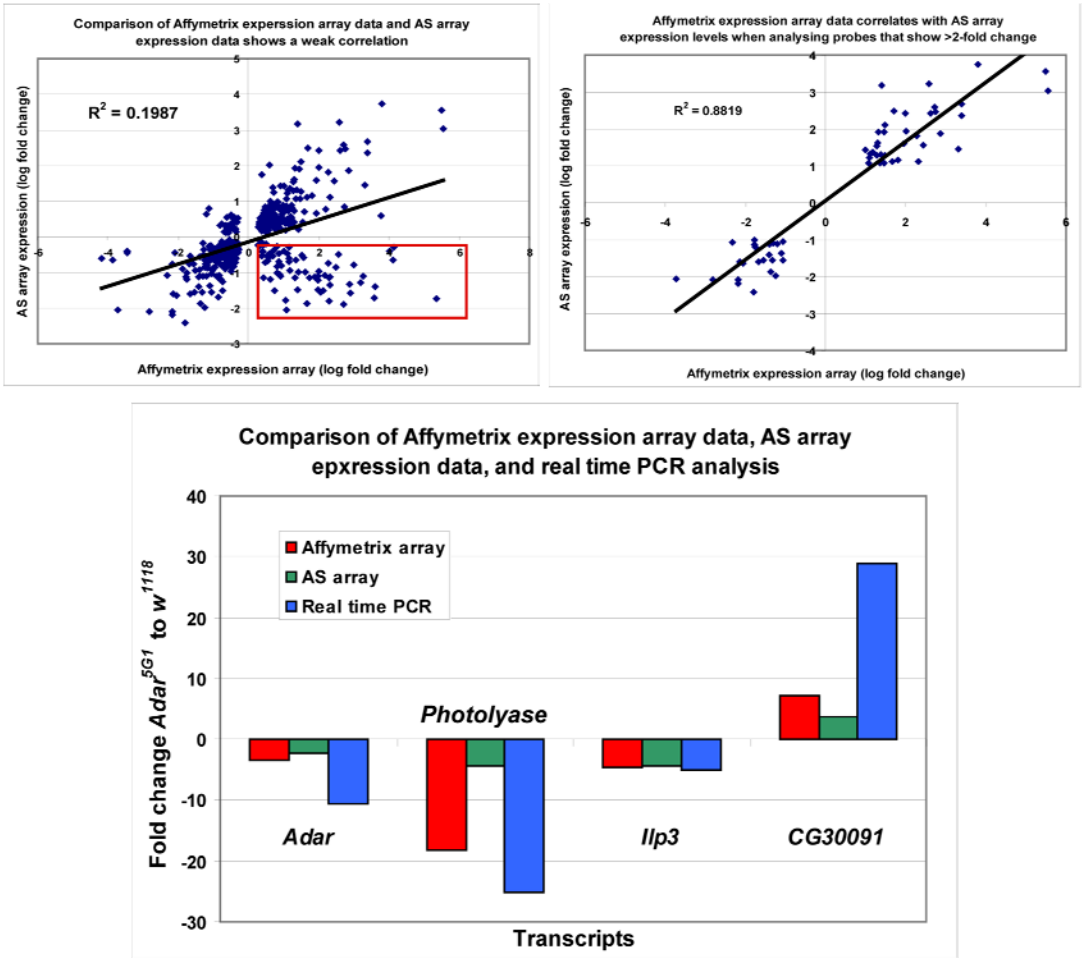


Figure 5.6: Comparison of array expression data
Comparison of Affymetrix expression array data with expression data from the AS array revealed a weak correlation (top left graph, $R^2 = 0.1987$). This was due to a large number of transcripts which showed increased levels according to the Affymetrix array but had decreased levels according to the expression array data obtained form the AS array (highlighted in red box). Limiting the analysis to transcripts which showed a >2-fold change in expression revealed a strong positive correlation (top right graph, $R^2 = 0.8819$), indicating that the significant changes in expression were detected by both arrays. Comparing four transcripts of interest across both microarrays and real time PCR analysis revealed the trends were the same however the fold change was usually low when analysed using the expression data from the AS array and high when analysed by real time PCR (*bottom panel*).

The AS array was performed to determine whether aberrant splicing was the cause of the altered transcript levels observed in *Adar*^{5G1} flies analysed by Affymetrix array. The expression data correlates well between the two arrays when analysing the most significant changes. However, very few transcripts which were altered according to the Affymetrix array showed altered splicing profiles. Of the 759 genes which

showed significantly altered junction use (p value of 0.001) only nine had altered expression according to the Affymetrix array. Of these seven were downregulated and two were upregulated according to the Affymetrix array. Therefore the different arrays analysed have identified two distinct subsets of genes, one group which is regulated at transcript level and the other which is regulated through changes in AS. This is consistent with other studies which have shown that distinct subsets of genes are regulated by AS changes and transcript level changes during the response to insulin treatment or stimulation of the *wingless* pathway in *Drosophila* S2 cells (Hartmann et al., 2009). In this study *Drosophila* S2 cells were treated with insulin and RNA was extracted 5 hours later. Using an AS array, changes in transcript level were detected in 149 genes and AS changes were found in 163 genes, following insulin treatment. In comparison the AS array on *Adar*-null flies identified many more transcripts which showed altered splicing, however the AS arrays used were different and the RNA used for analysing insulin-induced changes was isolated from a homogeneous population of *Drosophila* S2 cells rather than total fly head RNA. Distinct regulation of transcript level changes and AS changes has also been demonstrated during T-cell activation (Ip *et al.*, 2007). Here it was shown that genes from functionally distinct groups were regulated by changes in transcript levels and changes in AS events.

Confirmation of AS array changes by real time PCR

As there were many AS changes observed in the *Adar*-null flies further investigation focussed on the genes which showed the most significant changes (see Table 5.1). In a study on AS changes which occur during T-cell activation the authors indicated confirmation of AS array data is most reliable when restricted to the top third of the significantly altered genes (Ip et al., 2007). To confirm the changes observed at specific splicing junctions, primers for real time PCR were designed to amplify a short region which spans the affected junction. As a control primers were also designed to amplify a neighbouring constitutively expressed exon to analyse the local expression profile. Junction use was determined relative to the expression of the neighbouring exon to mirror the analysis performed on the AS array data. The AS

array analysis specifically highlights exon-exon junctions which are significantly altered compared to other junctions within the gene and the common exonic probes, whereas differences in expression between wild type and *Adar*-null flies are discounted. However, the expression of the local constitutive exon was also calculated relative to *GAPDH*.

Several genes from the list of the top 10 most significantly altered genes shown in Table 5.1 were investigated using real time PCR to confirm the changes observed on the AS array (see Figures 5.7 & 5.8). One of these genes, *Stretchin-Myosin light chain kinase* (*Strn-Mlck*) has a complicated locus resulting from the fusion of two different genes encoding *stretchin* and *myosin light chain kinase* (Champagne *et al.*, 2000). As well as transcripts which encode *Stretchin* and *Mlck* this locus also encodes a large fusion transcript which generates a large 926 kDa Titan-like protein. A recently discovered muscle protein termed Stretchin-klp has been mapped to the *Strn-Mlck* locus indicating the complex nature of the locus (Patel and Saide, 2005). There are two isoforms of Stretchin-klp, one of which is myosin dependent and expressed in the indirect flight muscles, the other is myosin-independent and is expressed in other tissues including the head. However *Stretchin-klp* is expressed from the centre region of the *Stretchin-mlck* gene and therefore does not contain the affected junctions.

Real time PCR analysis of an exon-exon junction (between exons 24 and 25) in *Strn-Mlck* which was significantly overrepresented in *Adar*-null flies confirmed the AS array data (Figure 5.7, junction 24-25 is indicated by a red arrowhead). Primers which span approximately 100bp across the junction indicate this region is overrepresented seven-fold in RNA extracted from *Adar*-null flies, and this is suppressed by expression of an *Adar* transgene in the cholinergic neurons. Analysis of a neighbouring exon which is expressed in the same transcripts that contain the exon 24-25 junction revealed that the expression of transcripts containing this region is downregulated approximately 10-fold in *Adar*-null flies. The difference in expression was partially rescued by expression of an *Adar* transgene in the cholinergic neurons. The real time PCR confirmed that the *Strn-Mlck* junction

between exons 24 and 25 is overrepresented in *Adar*-null flies. The *Strn-Mlck* locus is complicated therefore this result may indicate an alternative isoform is expressed from this locus in *Adar*-null flies.

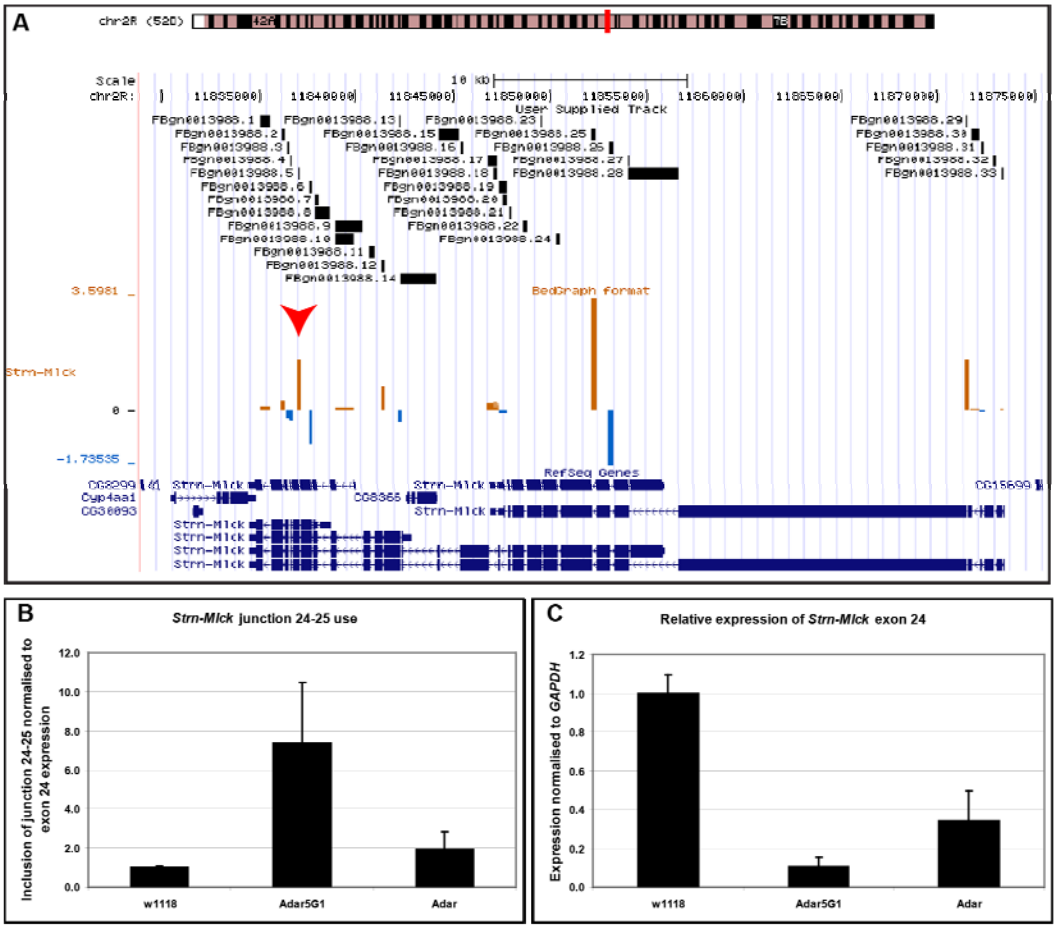


Figure 5.7: Real time PCR confirmation of increased junction use in the *Stretchin-Mlck* transcript.

Primers were designed which spanned the junction between exons 24 and 25 to analyse the level of junction use in w^{1118}, $Adar^{5G1}$ and $Adar^{5G1}$*:Cha-GAL4>UAS-Adar* flies. A) Junction 24-25 corresponds to probe set number 5 on the UCSC genome browser image (indicated with a red arrowhead). B) The junction between exon 24 and exon 25 was present at a higher level in $Adar^{5G1}$ than w^{1118}, and was restored to a lower level with expression of *Adar* in the cholinergic neurons. Expression was normalised to the expression of exon 24 to analyse the difference in junction use independent of differences in expression level. C) Primers were also used to investigate the expression level of exon 24, which was normalised to GAPDH levels. This revealed the transcript was expressed at a lower level in $Adar^{5G1}$ than w^{1118}, which was observed to a lesser extent with the AS array expression data. Probe set 25 was not investigated despite showing a greater change in inclusion as some data was missing from this junction and therefore the result was not significant (see Figure 5.3).

Graphing the AS array data from the *Dscam* locus revealed several junctions were altered in *Adar*-null flies, however some probes failed giving incomplete data sets for some of the junctions which means that statistical significance was not possible at these junctions. The majority of the exon junctions in *Dscam* were not significantly altered in *Adar*-null flies, which was a surprising result given that the mutually exclusive splicing of *Dscam* exon 6 cluster is regulated through formation of a stem-loop structure (Graveley, 2005). The stem-loop structure is formed between a "docking site" within the intron following the constitutively spliced exon 5 and a "selector site" upstream of each exon 6 alternatives, however the splicing factor hrp36 binds throughout the exon 6 cluster indicating perhaps ADAR would not have access to this stem-loop structure (Olson et al., 2007). Given the complex splicing patterns at the *Dscam* locus (described in the Introduction section on Alternative Splicing) this indicates ADAR does not play a role in the regulation of mutually exclusive splicing events within transcripts from the *Dscam* locus. However the single altered junction was investigated further as the transcripts derived from the *Dscam* locus have been extensively characterised.

One altered junction which showed significantly different expression in *Adar*-null flies was investigated further. The junction spanned the intron between exons 14 and 15, which are constitutively spliced exons. Real time PCR analysis of the region spanning this junction revealed that the junction was overrepresented in $Adar^{5G1}$ flies and this was rescued by expression of an *Adar* transgene in the cholinergic neurons (Figure 5.7). Analysis of a neighbouring exon revealed that transcripts containing this exon were also overexpressed in *Adar*-null flies. As this exon is constitutively expressed this indicates that there is higher expression of *Dscam* in *Adar*-null flies which is relieved by expression of *Adar* in the cholinergic neurons. However this change in expression levels is not dramatic and was not detected in the Affymetrix gene expression array data indicating it is not significant. PCR analysis of the region spanning the junction between exons 14 and 15 with primers located in exons 16 and 18 did not reveal a difference in products between wild type and *Adar*-null flies therefore a different approach to analyse all the transcripts from this region was used.

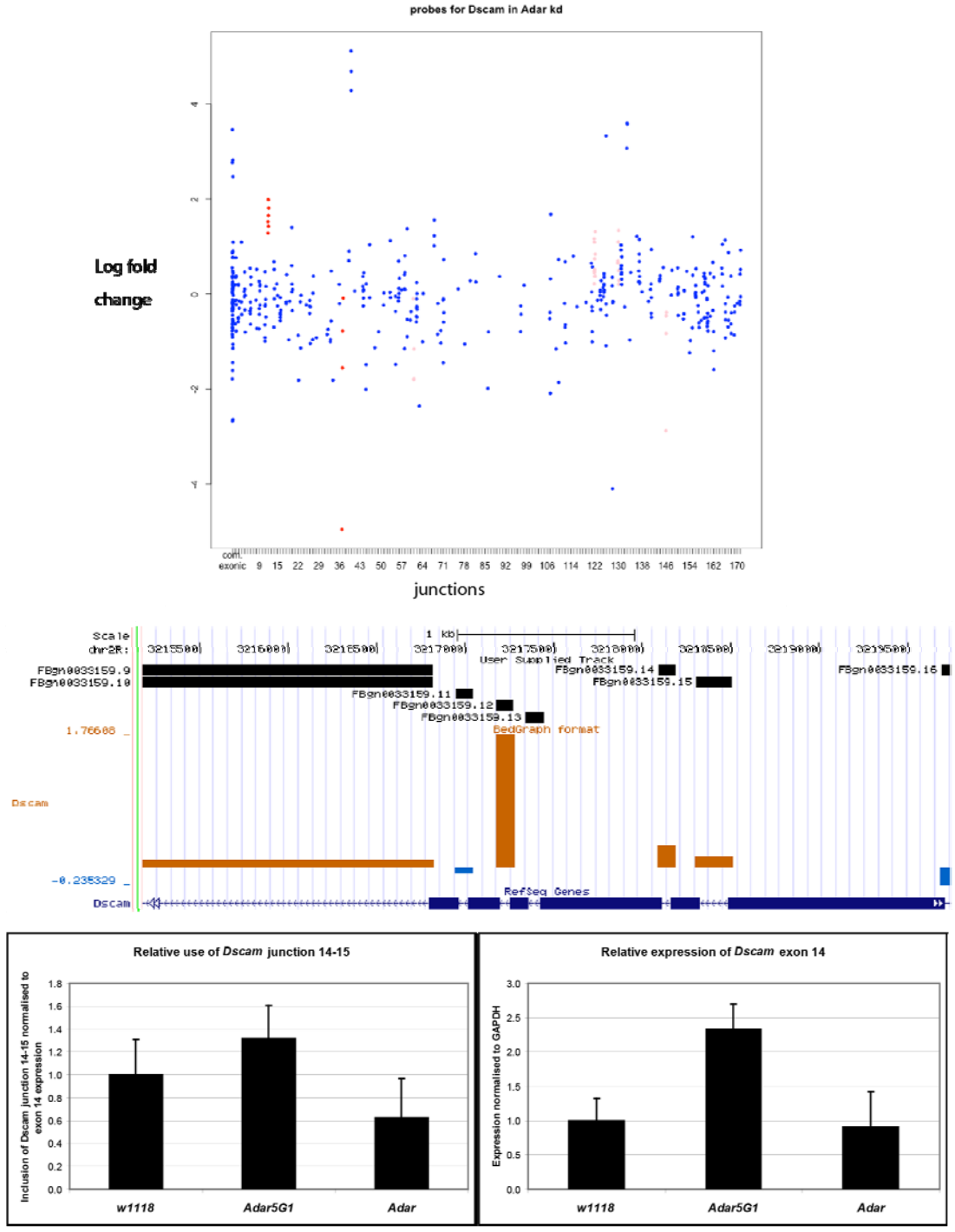


Figure 5.8: Altered junction use in *Dscam*

Surprisingly the majority of *Dscam* was spliced correctly in *Adar*-null flies (top panel). However, one junction (number 12) was significantly overrepresented in *Adar*-null flies (middle panel). This junction corresponds to the splice site between constitutively spliced exons 14 and 15. Real time PCR analysis using junction-spanning primers confirmed the AS array result showing that this junction is overrepresented in *Adar*-null flies (bottom left panel). Primers designed to a neighbouring exon to analyse the expression level of the *Dscam* transcript in this region showed that it was also elevated in *Adar*^{5G1} flies normalised to *GAPDH* expression, and this was rescued by expression of *Adar* in the cholinergic neurons.

RACE results for AS array targets

The AS array data revealed 759 genes showed significantly altered AS in *Adar*-null flies, and several of these changes were confirmed by real time PCR analysis of the regions containing the altered junctions. However, while the AS array indicates the affected genes it does not reveal how the transcripts from these genes are altered. Many genes which were altered according to the AS array showed differential use of single junctions, therefore to determine whether cryptic splice sites were utilised or processing of introns was altered 5' and 3' Rapid Amplification of cDNA Ends (RACE) was performed using the Firstchoice RLM RACE kit and protocol (Ambion). This technique allows amplification of mRNAs using one specific primer and a general primer directed either to the 5' cap structure or to the poly A+ tail. This method was chosen as it should reveal all the isoforms transcribed from one gene which contain the regions targeted by the gene-specific primer and extend to either the 5' or 3' end of the transcript. 5' and 3' RACE was performed using 1.5µg of total RNA isolated from w^{1118}, $Adar^{5G1}$ and $Adar^{5G1}$*:Cha-GAL4>UAS-dAdar* flies using gene specific nested PCR primers. Two rounds of amplification of 30 cycles were used with 2ul of the first PCR product used as template for the second reaction. RACE products were analysed on 2% agarose gels and bands were purified and cloned into T-easy for sequence analysis.

Unfortunately 5' RACE was not successful and no specific bands were observed following gel electrophoresis; this was likely to be due to inefficient ligation of the 5' linker prior to cDNA amplification. 3' RACE was performed with gene-specific primers for several transcripts identified from the AS array in the attempt to discover whether cryptic splice sites were utilised in the absence of ADAR. 3' RACE analysis using gene-specific primers for the *Dscam* locus (Figure 5.9) and the *dumpy* locus (Figure 5.10) revealed different products were present in RNA isolated from *Adar*-null flies when compared to wild type RNA.

Surprisingly sequence analysis of the 3'RACE products from *Adar*-null flies did not reveal retained introns or use of cryptic splice sites. Instead these products showed premature polyadenylation of transcripts from both the *Dscam* locus and the *dumpy*

locus (See Figures 5.9 & 5.10). This result was not expected, however these products were observed at both loci and were not present in w^{1118} or *Adar*^{5G1}*:Cha-GAL4>UAS-dAdar* RNA.

Analysis of 3' RACE products from the *Dscam* locus using nested gene specific forward primers in exon 14 revealed a lower molecular weight band of approximately 100bp in the *Adar*^{5G1} sample that is not present in w^{1118} or *Adar*^{5G1}*:Cha-GAL4>UAS-dAdar* RNA (Figure 5.9). A higher molecular weight product of approximately 400bp is present in all three RNA samples. A control of w^{1118} RNA which was incubated without reverse transcriptase enzyme was included to ensure no DNA contamination was present and no PCR product was detected in this sample. Sequence analysis of the 3' RACE products revealed that the ~400bp product corresponds to the *Dscam* transcript exons 14, 15, 16, and 17 with part of the intronic region between exons 17 and 18. This is due to a large run of adenosine residues that occurs within the intron between exons 17 and 18, which would result in amplification of this region during the first strand synthesis reaction with an oligo dT primer. This product occurs in all three RNA samples and therefore provides an internal control to ensure that *Dscam* transcripts are present and amplified in all RNA samples. The 3'RACE product of approximately 100bp which was only amplified from *Adar*^{5G1} RNA corresponds to correctly spliced *Dscam* exons 14 and 15, but terminates within exon 15 with a run of adenosine residues indicative of polyadenylation.

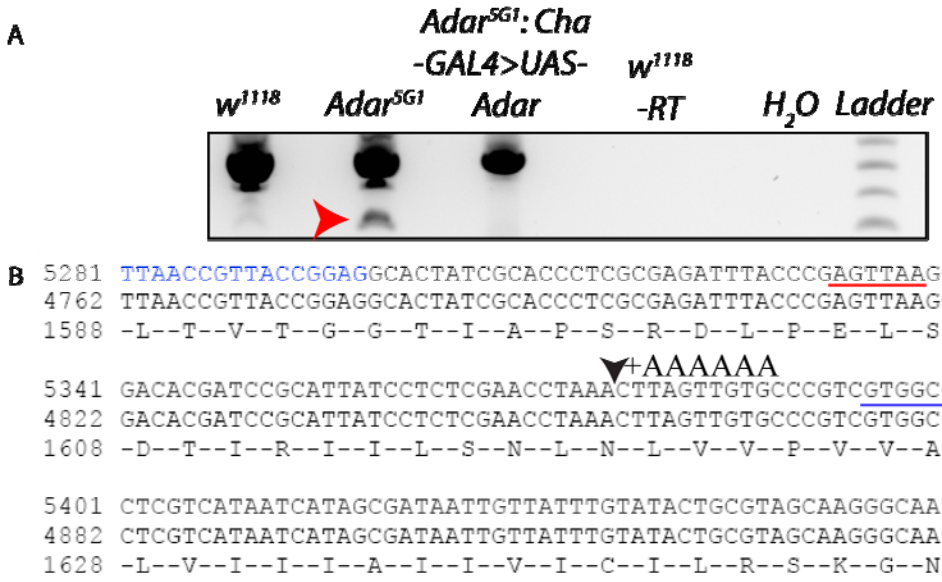


Figure 5.9: *Dscam* 3' RACE

(A) 3' RACE was performed with nested forward primers which are specific for the *Dscam* transcript and occurred in a common exon (exon 14). 3' RACE products were electrophoresed on a 1.5% agarose gel, which revealed an extra band that was present in *Adar*-null flies that was not seen in controls. (B) Sequence analysis of the cloned 3' RACE products revealed that there was a premature polyadenylation site in transcripts from $Adar^{5G1}$ flies (indicated by red arrowhead in A). This was not observed in RACE products cloned from w^{1118} or flies expressing *Adar* in the cholinergic neurons. The intense band present in all samples correponds to correctly spliced *Dscam* transcript containing exons 14, 15, 16, and 17. This product terminates within the intron between exons 17 and 18 where there is a long run of adenosine residues which mimics a polyadenylation sequence and was therefore amplified with the oligo dT primers in the first strand synthesis reaction. Analysis of the surrounding sequence indicated putative polyadenylation signals (underlined in red) were present although the canonical AAUAAA was not observed. A U/GU-rich downstream sequence was also identified (underlined in blue, where T=U as the DNA sequence is shown).

Similarly analysis of 3'RACE products using gene specific forward primers within *dumpy* exon 33 revealed extra products were present in the $Adar^{5G1}$ RNA sample that were not observed in w^{1118} or $Adar^{5G1}$*:Cha-GAL4>UAS-dAdar* RNA (Figure 5.10). A single band of approximately 650bp was present in all three RNA samples; sequence analysis revealed this was not from the *dumpy* locus but was a contaminating product amplified from a different locus. However, this was consistently amplified in all samples. Two extra bands present in the $Adar^{5G1}$ RNA, one of approximately 550bp and the other of approximately 850bp, correspond to correctly spliced product from dumpy exons 33 and 34. However theses products both terminate within exon 34 of the dumpy locus with a run of adenosine residues which are not present in the annotated genomic sequence from this region indicative of alternative polyadenylation.

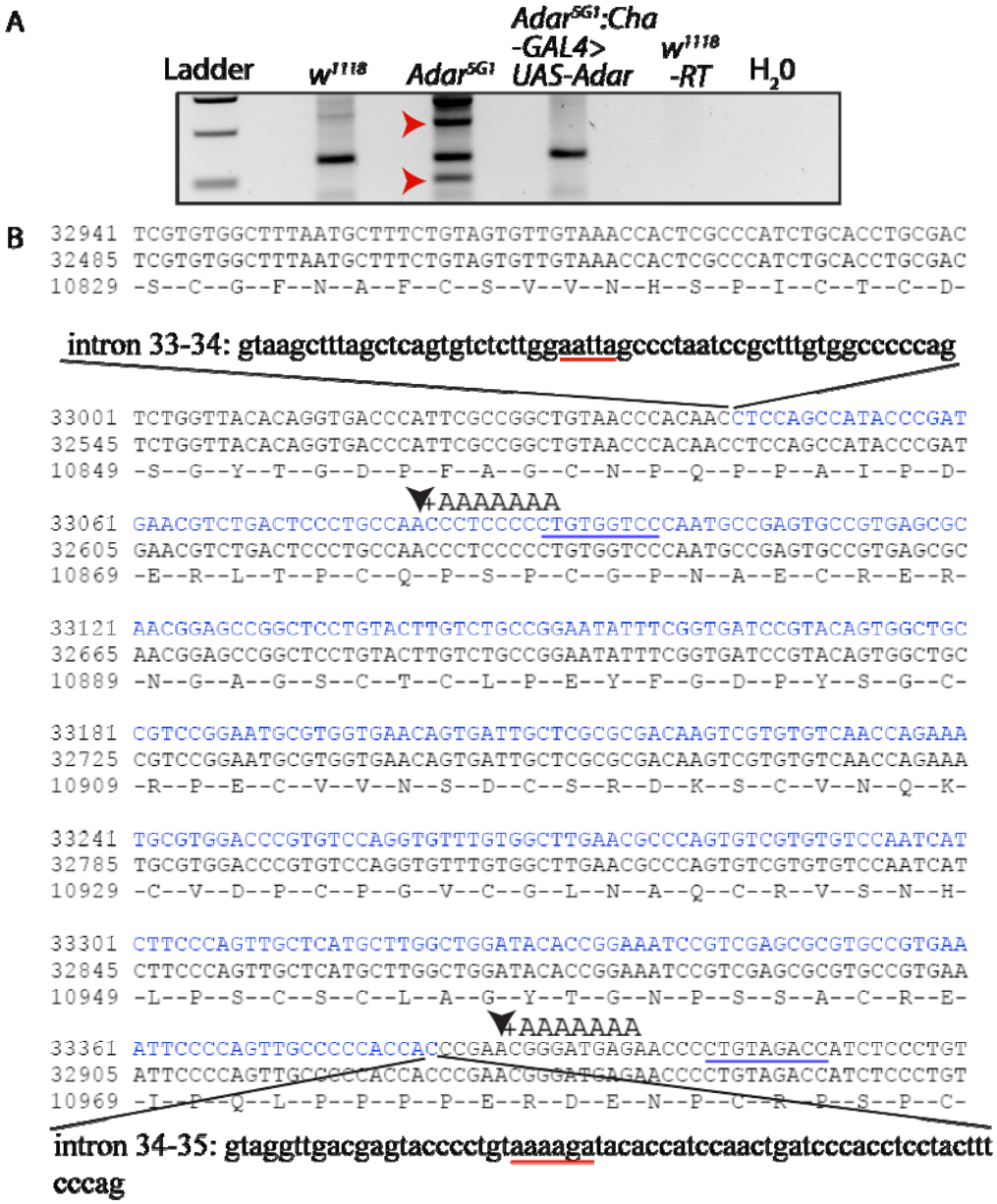


Figure 5.10: *Dumpy* 3' RACE

(A) 3' RACE was performed on RNA extracted from w^{1118}, $Adar^{5G1}$, and $Adar^{5G1}$:*Cha-GAL4>UAS-Adar* flies to determine whether the altered splicing patterns observed were the result of use of cryptic splice sites or intron retention. 3' RACE was performed with nested forward primers specific to the *dumpy* locus and reverse primers supplied with the Firstchoice RLM RACE kit (Ambion). Analysis of 3' RACE products by gel electrophoresis revealed multiple bands in the $Adar^{5G1}$ RNA that were not present in the w^{1118} or $Adar^{5G1}$:*Cha-GAL4>UAS-Adar* RNA (B) Sequence analysis of the 3' RACE products cloned from $Adar^{5G1}$ flies showed premature polyadenylation at two sites within *dumpy* exon 34. No obvious alternative polyadenlyation signal was found in the coding sequence however both sites occurred close to a splice junction and examination of the intronic sequence revealed putative alternative polyadenylation signals (underlined in red) although these differ from the canonical AAUAAA signal. The sequence downstream of the alternative polyadenylation sites was U/GU-rich (underlined in blue) indicating these regions could be binding sites for the polyadenylation cleavage factors.

Canonical polyadenylation occurs at specific locations within transcripts directed by conserved *cis* element consensus sequences, the process of 3' end processing is reviewed in (Mandel *et al.*, 2008). Briefly, a polyadenylation signal (PAS) sequence of AAUAAA site which is located 10-35bp upstream of the cleavage site is required for polyadenylation. The cleavage site itself which is not highly conserved and often has a cytosine and an adenosine residue immediately preceding it. Following this there is a downstream sequence element (DSE) comprising U/GU-rich sequence which is located up to 70bp downstream of the polyadenylation site and provides a binding site for the Cleavage Stimulation Factor (CstF). Many protein factors are involved in cleavage and polyadenylation (Mandel et al., 2008), and these processes have been linked to transcription and alternative splicing, which is reviewed in (Proudfoot, 2004).

However, examples of polyadenylation are increasingly found which do not contain these canonical sequences. Polyadenylation can occur at regions that show variant sequences which resemble the canonical sequences but are not exact matches. One study found that cleavage and polyadenylation in *Drosophila* ESTs occurs in 22.48% of sequences analysed in the absence of a recognisable polyadenylation signal (Retelska *et al.*, 2006). Further this study showed approximately half of the ESTs analysed contained the canonical AAUAAA consensus sequence indicating variation is widespread.

Analysis of the sequences upstream of the alternative polyadenylation sites found in *Adar*-null flies revealed that the consensus polyadenylation signal AAUAAA was not present. However, sequences which closely resemble the canonical site were present in *Dscam* an AGUUAA sequence is present 40bp upstream of the polyA site (Figure 5.9). The *dumpy* transcript did not contain an obvious polyadenylation signal, however a less similar sequence of AAUUA was present near the first site and an AAAAGA was found near to the second site (Figure 5.10).

Importantly the regions surrounding the alternative polyadenylation sites do not contain multiple adenosine residues indicating amplification of these sequences is

not likely to be due to spurious binding of the oligo dT primers during cDNA amplification. Further, in the *Dscam* analysis, this type of spurious priming does occur in all the RNA samples analysed and is observed as a strong band of higher molecular weight than the alternative polyadenylated transcript. This is clearly present to approximately equal amounts in all three RNA samples indicating the prematurely polyadenylated product amplified from $Adar^{5G1}$ RNA only accounts for a small fraction of the product from this locus.

Cleavage and polyadenylation of nascent transcripts normally occurs downstream of a translation stop codon, however only the 3' regions of the *Dscam* and *dumpy* sequences were analysed here. Further investigation of the full transcripts is required to determine whether these sequences arise from splice variants which terminate prior to the polyadenylation sites identified. Both of these genes encode long transcripts which makes amplifying and cloning all the full length variants from one locus difficult.

In these experiments total RNA was used rather than polyA+ purified RNA, either of which can be used according to the manufacturer's protocol. Total RNA was used as starting amounts of fly head RNA were limiting and when RNA amounts are low the loss during polyA+ purification can result in insufficient starting material. Further, the aim of these experiments was to characterise changes observed with the AS array and this was performed on oligo dT amplified total RNA. However further analysis using polyA+ purified RNA as starting material is necessary to exclude partially processed transcripts.

These results were not expected. Several other transcripts identified as altered according to the AS array were investigated using 3' RACE and multiple products were obtained. However sequence analysis revealed no difference in products from wild type and *Adar*-null flies. Multiple cycles of amplification are required to detect these products and partially processed products may be amplified. However, the *Dscam* region was amplified from all three RNA samples analysed indicating the 3' RACE protocol used was able to amplify this region from $Adar^{5G1}$ RNA. One

explanation for these results could be altered secondary structure of the RNA isolated from *Adar*^{5G1} flies. The total RNA was denatured at 70^{o}C prior to first strand cDNA synthesis which makes this unlikely however further work is required to exclude this possibility. Another explanation for these results is that these transcripts are aberrantly spliced such that a region which contains a run of adenosine residues is spliced with the region amplified using the gene-specific primer. The *Dscam* locus contains a run of adenosines in the intron between exons 17 and 18, if aberrant splicing occurred this may appear as a polyadenosine tail in the mature transcript. However it would be difficult to distinguish these two products using RACE. Further investigation of all transcripts from the *Dscam* locus is required.

Is alternative polyadenylation widespread in Adar^{5G1} flies?

The discovery of two genes which appear to show alternative polyadenylation in *Adar*-null flies led to the question of whether ADAR influences polyadenylation site choice in other genes. If the alternative polyadenylation site occurs upstream of the canonical site, as with the two examples shown, then this could lead to truncated transcripts which do not contain the 3'UTR region to which the Affymetrix array probes are designed. Initial investigation of genes which were downregulated in *Adar*-null flies according to the Affymetrix array and also showed aberrant splicing according to the AS array revealed that there were 7 genes which fitted this criteria. Examination of the AS data for these genes indicated two of the seven showed decreased use of splice junctions at the 3' end of the gene (see Figure 5.10). Both *CG11474* and *Nocturnin* genes showed AS array data profiles which could be consistent with alternative polyadenylation in *Adar*-null flies leading to truncated transcripts lacking the 3'UTR region used to analyse expression data on the Affymetrix array. 3' RACE analysis to determine the sequence prior to the polyadenylation site is required to examine this hypothesis further.

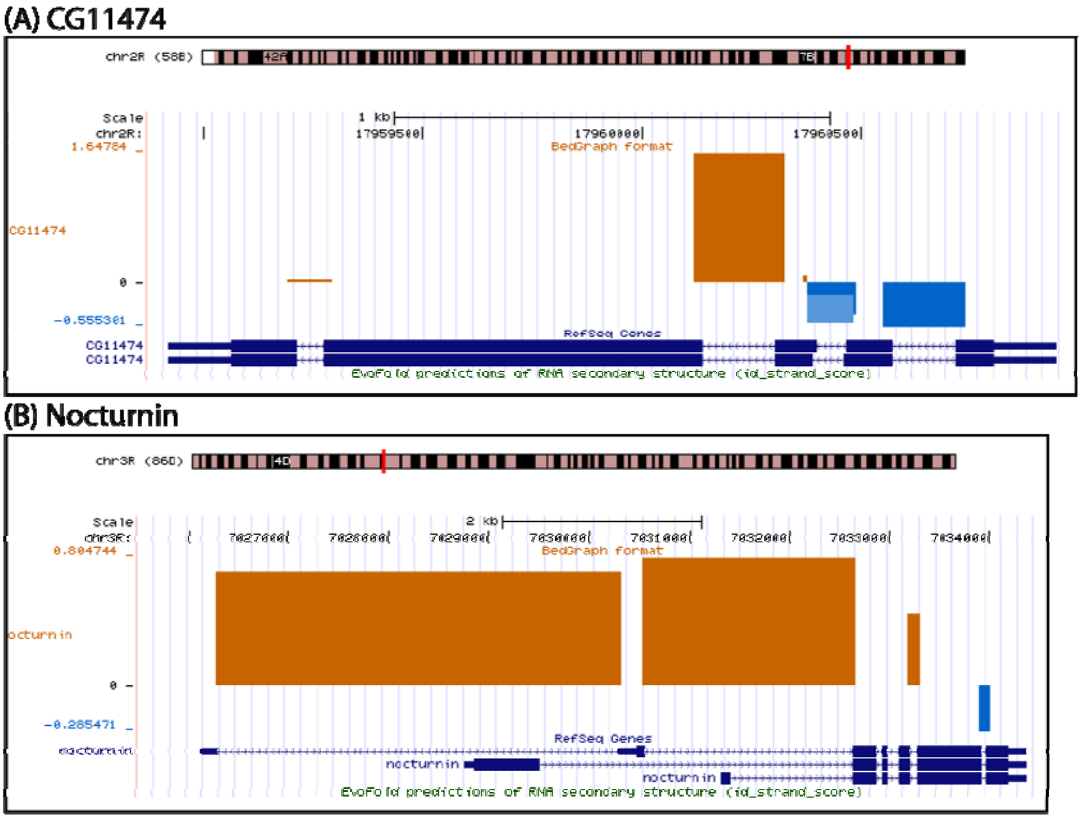


Figure 5.11: Candidates for alternative polyadenylation in *Adar*-null flies
Following the 3' RACE results for *Dscam* and *dumpy* transcripts analysis of AS array data indicated that two other transcripts are candidates for alternative polyadenylation site use. *Nocturnin* and *CG11474* were identified as downregulated by the Affymetrix array data (*Nocturnin* -1.44-fold downregulated and *CG11474* -1.79-fold downregulated) and show variation in AS across the gene. Intriguingly the AS data for both genes shows decreased junction use close to the 3'UTR region where the Affymetrix array probes hybridise to. This could indicate an alternative polyadenylation site is used in *Adar*-null flies which creates transcripts lacking the canonical 3'UTR region, however further work is required to characterise the transcripts associated with these genes.

B52/Srp55 expression is not altered in Adar^{5G1} flies

Numerous splicing changes were detected in *Adar*-null flies however one of the genes affected encodes a splicing factor B52 (Figure 5.12 A). B52 is the *Drosophila melanogaster* homolog of the mammalian splicing factor Srp55 and it is the predominant SR protein expressed in *Drosophila* brain (Hoffman and Lis, 2000). The overrepresented splice junction in B52 leads to inclusion of a large exon generating a long 3'UTR region and may therefore affect mRNA stability or translation. However, the junctions downstream of the overrepresented junction do not show the same pattern; in fact the splice junctions in the 3' region of the gene are all underrepresented in *Adar*-null flies. This may indicate another isoform is present which stops upstream of the known isoforms or it may indicate that this gene undergoes alternative polyadenylation as demonstrated for the *Dscam* and *dumpy* genes. The AS profile of B52 is similar to those of *Nocturnin* and *CG11474* shown in Figure 5.11 indicating these genes may undergo alternative polyadenylation in flies which lack the *Adar* gene.

B52 was one of the genes affected in *Adar*^{5G1} flies according to the AS array data and an antibody was available, therefore protein levels were analysed to determine whether a difference in B52 could explain the observed changes in splicing. The anti-SR protein antibody 1H4 (Invitrogen) detects numerous SR proteins in mammalian cell extract however it has been shown to be specific for B52 in *Drosophila* S2 cells by siRNA depletion of B52 mRNA (Marcucci et al., 2009). The 1H4 epitope is phospho-dependent therefore protein extracts from fly heads were prepared in the presence of phosphatase inhibitors. A band of approximately 70kDa corresponding to the B52 protein was detected in all strains analysed, although the protein migrated at a higher molecular weight than expected, likely due to the known phosphorylation events which regulate activity (Figure 5.12 B). Slightly more protein was observed in the *Adar*^{5G1} sample; however examination of the ponceau stain to reveal the total protein loaded indicated that there was slightly more protein loaded in the *Adar*^{5G1} lane (Figure 5.12 B; lane 2). A control of HEK293T extract was included to ensure the antibody detected multiple phosphorylated SR proteins in mammalian extract. A HEK293T extract which has been de-phosphorylated with

calf intestinal phosphatase (CIP) prior to loading was also included to show the phospho-specificity of the 1H4 antibody.

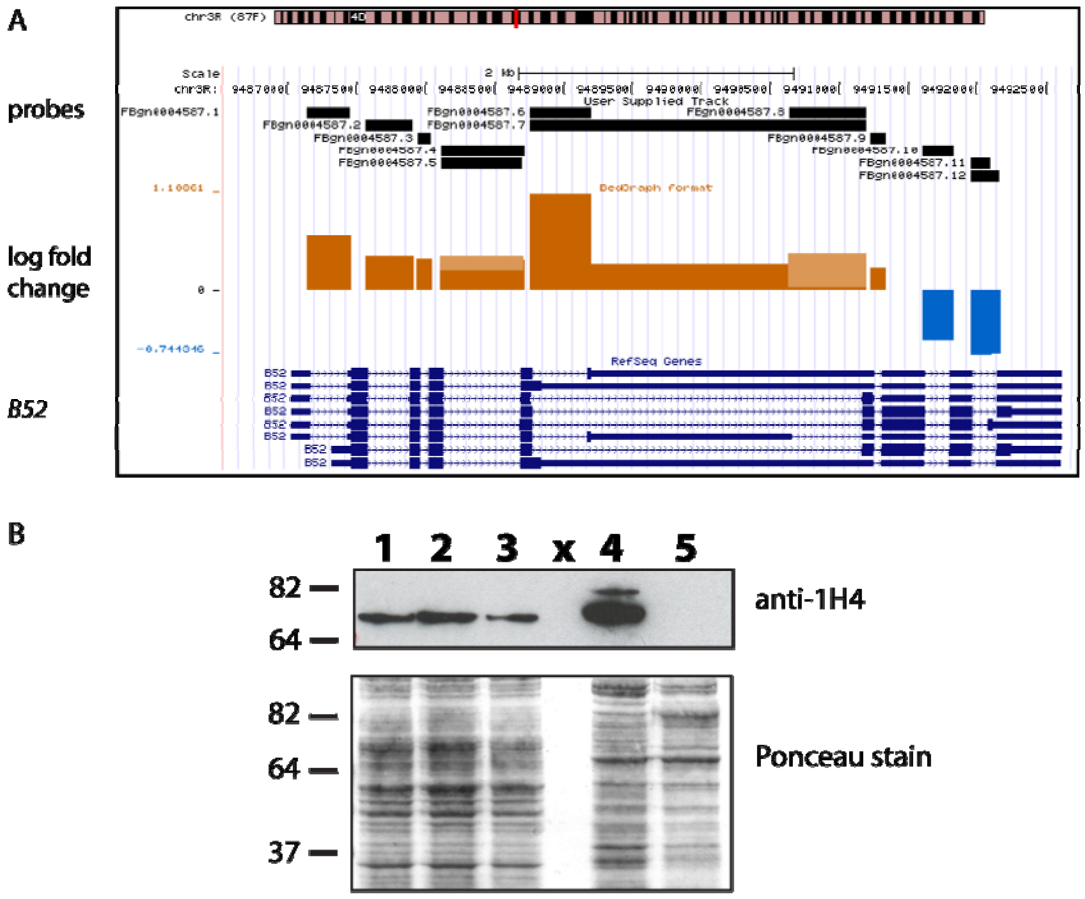


Figure 5.12: B52/Srp55 protein levels are not altered in *Adar*^{5G1} flies.
A) In the B52 gene one splice junction was significantly altered in *Adar*-null flies. The overrepresented junction leads to inclusion of a large exon within the 3'UTR and is present in two out of eight annotated transcripts from the *B52* locus. To determine whether the alteration in 3'UTR length affected the level of active B52 protein an immunoblot was performed with protein extract from fly heads. B) Immunoblot with an antibody which recognises a phospho-epitope found on many SR proteins (Invitrogen) on fly head extract from 1) *w*^{1118}, 2) *Adar*^{5G1}, 3) *Adar*^{5G1}*:Cha-GAL4>UAS-Adar*, 4) HEK293T cell extract and 5) HEK293T extract which was pre-treated with phosphatase to remove phosphorylation marks. Lane x is empty. The immunoblot (top) shows a single band in *Drosophila* head extracts (lanes 1-3) which corresponds to phosphorylated B52 protein. Multiple bands were present in the HEK293T positive control (lane 4), which disappeared following phosphatase treatment (lane 5). A scan of the ponceau stain is included to show that the difference observed in B52 levels is probably due to slight differences in the amount of protein loaded.

5.3 Discussion

Analysis of RNA from flies lacking the *Adar* gene revealed 759 genes which showed significantly altered splice junction use. This is a higher number than was expected, however many of these genes showed altered use of one junction rather than a coordinated change in isoforms. When compared to other AS array data from *Drosophila* our data indicated a larger number of affected genes. However, previous studies have used cell lines depleted for specific splicing factors (Blanchette *et al.*, 2009) or following stimulation of specific pathways (Hartmann et al., 2009). Analysis of total head RNA includes many different cell types and provides a global analysis of the splicing status, but the analysis cannot distinguish signals from different splice variants which contain the same splice junctions. However, AS arrays have been used successfully to determine the substrates of the brain specific NOVA proteins indicating global analysis of samples which contain multiple different cell types can be used to identify specific substrates (Ule *et al.*, 2005).

Detailed investigation of transcripts from two genes which are altered in *Adar*-null flies did not reveal multiple alternatively spliced isoforms. Instead transcripts were detected with abnormal polyadenylation, which was an unexpected result. These results were obtained using the 3'RACE technique that includes a high number of amplification cycles which increases the chances of amplifying spurious products; however this is unlikely for several reasons. Firstly, the alternative polyadenylation sites did not occur in regions that were rich in adenosine residues which may hybridise to oligo dT primers. Secondly the wild type or canonical sequence was amplified from all three RNA samples indicating these alternative polyadenylation sites are only utilised in the absence of ADAR and only contribute a small fraction of the total transcripts from these loci. Thirdly, the samples were denatured at 70^{o}C prior to first strand cDNA synthesis to melt any secondary structure therefore it is unlikely that altered structure in the RNA from flies lacking the *Adar* gene could give rise to the observed results. However these results need to be verified using polyA purified RNA as starting material to ensure partially processed transcripts are excluded.

Analysis of AS array data in the light of these findings has identified several genes that may produce transcripts which are alternatively polyadenylated in *Adar*-null flies indicating this may be a widespread phenomena. However further work is required to determine the extent of this alternative polyadenylation and the role that ADAR plays in canonical polyadenylation in these transcripts.

There are several mechanisms which could explain the alternative polyadenylation observed in *Adar*-null flies. Firstly, ADAR could act to promote canonical polyadenylation in wild type flies through interactions with the 3' region of nascent transcripts. Several papers have described RNA editing in the 3'UTR region of transcripts isolated from *Caenorhabditis elegans* and humans indicating ADAR binds to the 3'UTR region of some transcripts (Morse *et al.*, 2002; Morse and Bass, 1999). Indeed the majority (77%) of the new editing events described by Li *et al*, occur within non-coding regions (Li et al., 2009). In addition, several *Drosophila* ADAR substrates identified by Stapleton *et al*, (2006) show editing in the 3'UTR region indicating these may be informative for future analysis (Stapleton et al., 2006). The majority of miRNA target sites are located within 3'UTR regions and use of alternative polyadenylation sites could alter the regulation of the transcript by removing miRNA target sites.

Secondly, ADAR could bind to the regions identified and block these alternative polyadenylation sites in the wild type fly brain. Therefore in the absence of ADAR these sequences would be exposed and available to interact with the 3' end processing machinery. Thirdly, ADAR could influence the secondary structure of these transcripts. It has been demonstrated that absence of the RNA helicase encoded by *maleless* can result in aberrant editing and splicing catastrophe for the *para* transcript (Reenan et al., 2000). This indicates that the secondary structure which is required for editing by ADAR is complex and can inhibit splicing in the absence of a RNA helicase to help resolve it. A highly structured 3' end of transcripts may inhibit normal processing by cleavage and polyadenylation factors. Fourthly, it cannot be excluded that these alternatively polyadenylated transcripts occur in wild type flies but are rapidly degraded. ADAR could function in an RNA

surveillance role to aid in the degradation of aberrantly polyadenylated transcripts, and therefore these transcripts are stabilised and detected in the absence of ADAR.
Lastly, it cannot be ruled out that the observed effects may be indirect. ADAR may edit and regulate a key component in the polyadenylation machinery, and in the absence of this editing event polyadenylation site choice may be de-regulated. In support of this hypothesis one gene which showed altered AS in *Adar*-null flies, *CG13900*, encodes a component of the U2 snRNP which participates in the splicing reaction (Mount and Salz, 2000). However CG13900 also contains an Interpro domain which is also found in the cleavage and polyadenylation stimulation factor (CPSF) indicating alterations in this transcript may affect the processing of other transcripts. Expression of *CG13900* was not significantly altered in *Adar*-null flies which suggests if it is miss-regulated it is not at the transcript expression level. However, the altered splice junction use occurred in the 3'UTR region and could therefore affect translational regulation. Further work to determine the level of CG13900 protein in *Adar*-null flies is required as it has been shown that the concentration of cleavage and polyadenylation factors can influence polyadenylation site choice.

The *Drosophila* Sex-lethal RNA-binding protein regulates female specific expression of a polyadenylation variant of enhancer of rudimentary gene through binding to the G/U-rich region and competing with CstF-64 (Gawande *et al.*, 2006). This indicates that alternative polyadenylation sites are used in *Drosophila* and RNA-binding proteins can influence the choice of polyadenylation site through preventing access to cleavage and polyadenylation factors. This mechanism may be used more widely than has previously been thought.

Interaction between RNA editing and polyadenylation is not a new phenomenon as an example in polyomavirus has been described. Editing of the first adenosine residue of the canonical AAUAAA polyadenylation signal in late viral transcripts is thought to decrease use of this polyadenylation signal (Gu et al., 2009). This allows run-through of the transcription termination signals and generates multigenomic

primary transcripts which are then processed by alternative splicing to produce multiple copies of late viral transcripts.

This chapter details the beginning of an investigation into the role of ADAR in alternative splicing. A subset of transcripts have been identified which show altered splicing in the absence of ADAR, however further work is required to characterise the role of ADAR in these splicing events. Crucial to the investigation of the role of ADAR in AS is determining whether ADAR binds to these transcripts in the regions which show altered splicing profiles. Binding of ADAR to transcripts which show altered splicing was to be analysed using the RNA immunoprecipitation protocol described in Chapter 4 however technical difficulties with this method prevented further characterisation of the interactions between ADAR and the identified transcripts.

A previous study into the role of *Drosophila* hnRNP family members in splicing utilised an AS array in conjunction with RNA SELEX to successfully identify consensus sequences for these RNA binding proteins (Blanchette et al., 2009). However SELEX with ADAR2 revealed that ADARs do not bind to a specific consensus sequence; it is the secondary structure which determines ADAR-binding (Hallegger et al., 2006). AS array profiles from known substrates of ADAR indicate aberrant splicing occurs in regions which neighbour ADAR binding sites. Therefore a particular focus on the local secondary structure surrounding junctions which show altered splicing according to the AS may be informative. However, complicated secondary structures including stem loop structures with over 1.5kb of looped RNA and complex pseudoknots have been described for known substrates of *Drosophila* ADAR which makes modelling the secondary structure of some transcripts difficult (Reenan, 2005). Detailed bioinformatic analysis of the regions which show altered splicing in *Adar*-null flies may be informative.

Microarray analysis of transcript changes which occur during mammalian T cell activation revealed one group of transcripts were regulated through alternative splicing and another distinct group were regulated at the transcript level (Ip et al.,

2007). This is similar to what we observed in *Drosophila* lacking the *Adar* gene where there is very little overlap between genes which are altered according to the Affymetrix array and genes which show aberrant splicing according to the AS array. Interestingly the two groups of genes described by Ip et al (2007) were functionally distinct, indicating regulation of specific pathways or processes may occur through either regulation of transcript level or AS events (Ip et al., 2007). A similar effect was observed in *Drosophila* S2 cells when analysing the response to insulin treatment or activation of the *wingless* pathway with functionally distinct groups of genes regulated either at the transcript level or through AS (Hartmann et al., 2009).

GO term analysis performed on the datasets obtained from the Affymetrix expression array showed that cytochrome genes were miss-regulated at the gene expression level indicative of oxidative stress response. Whereas GO term analysis on the AS array data revealed a different set of genes displayed altered splicing in *Adar*-null flies. These included several genes which contain a tropomyosin domain and are involved in movement and are known to be extensively alternatively spliced in *Drosophila*. Glutamate receptor subunit genes were also overrepresented in the AS array altered genes which may indicate a role for ADAR in processing these transcripts. This would suggest evolutionary conservation of ADAR substrates as several mammalian glutamate receptor subunit transcripts are known substrates of ADAR. Mammalian ADARs edit several ion channel receptor subunit transcripts which alters the processing of these subunits into functional receptors (reviewed in (Daniel and Ohman, 2009); a similar mechanism may occur in *Drosophila*.

AS arrays provide a valuable tool for examining the global splicing profile however they are dependent on a well annotated genome and it is likely there are more transcript variants expressed than are detailed on the AS array as novel isoforms are regularly discovered. Numerous AS changes were observed in RNA isolated from *Adar*-null flies indicating ADAR may have an additional function as an RNA-binding protein. The nature of the splicing changes, where single splice junctions were often dramatically affected in isolation could indicate that there are large

numbers of unannotated cryptic splice sites which occur in previously uncharacterised isoforms. However RACE analysis of several genes which were altered in *Adar*-null flies did not reveal retained introns or cryptic splice site use. This suggests that the single affected junctions may be bound by ADAR in wild type flies and therefore processing of these junctions is altered in *Adar*-null flies. Further work is required to determine whether ADAR binds to transcripts in the region of aberrant splicing.

Chapter 6: Discussion and Future Work

The discovery of a putative novel member of the AID/APOBEC family of cytidine deaminases (Rogozin et al., 2005) that is expressed in the testis and which may have a role in the innate immune system led to many questions. The aim of this thesis was to characterise the biochemical properties and expression pattern of this putative novel deaminase. The primary question to answer was whether A4 is a catalytically active deaminase and what substrate specificity it has. A4 is conserved across species from *Xenopus tropicalis* to human indicating it has a biological function (Rogozin et al., 2005). Furthermore, key residues within the catalytic domain are also conserved so presumably a zinc atom is coordinated within the active site, which is a shared feature of the cytidine deaminase family. This is suggestive that A4 has a catalytic deaminase function, however other residues which are conserved across other AID/APOBEC family members are not present which could indicate different substrate specificity.

Despite generation of several different A4 expression vectors and the use of three different expression systems, recombinant A4 could not be purified in sufficient quantities for biochemical characterisation due to its inherent insolubility. Other members of the AID/APOBEC family are also difficult to purify due to insolubility (Smith, 2007), however the presence of a coiled-coil domain in the carboxy terminus of A4 seems to have exacerbated the problem. Coiled-coil domains are involved in protein-protein interactions and this may enhance the aggregation of A4 when overexpressed for purification. Removal of the coiled-coil domain may improve the solubility of A4 for purification; however this may affect the behaviour of the protein as it could lose the ability to interact with a co-factor.

An alternative approach for the biochemical characterisation of A4 is to purify endogenous A4 using classical size exclusion chromatography methods. This may overcome the solubility problem however as there are problems detecting the endogenous protein it is difficult to predict whether testis or sperm should be used as starting material. Therefore it may not be possible to perform *in vitro* biochemical analysis on A4. Indeed, other members of the AID/APOBEC family have been analysed in cell culture systems or using the *E.coli* DNA mutator assay to circumvent

the need for purified protein preparations. The lack of a suitable germline-derived cell line precludes analysis of A4 in a cell culture model which would re-create the endogenous A4 environment; however expression in a heterologous cell culture line may prove useful when the substrate of A4 has been identified.

The insolubility of A4 could also have contributed to the lack of DNA deaminase activity in the *E.coli* DNA mutator assay as insoluble aggregated A4 may not have access to the DNA. However, this may be an accurate result and indicate that A4 does not have catalytic deaminase activity on DNA. Independent verification of the substrate specificity of A4 is required yet without the ability to purify recombinant A4 this may prove difficult.

Apobec4 transcripts are expressed at the round spermatid stage of spermatogenesis, which is a highly specialised cell type. However during this stage of spermatogenesis mRNAs encoding proteins which are required for spermiogenesis and fertilisation are also transcribed and stored in translationally silent states for use after chromatin compaction when DNA cannot be accessed. Therefore detection of *Apobec4* mRNA is not indicative of A4 protein expression. The discovery of a target site for miRNA-741 in the *Apobec4* 3'UTR indicates expression of A4 protein may be regulated post-transcriptionally. miRNA-741 shows a similar expression pattern to *Apobec4* suggesting they are expressed at the same stage of spermatogenesis. Further experiments are required to determine whether they are expressed in the same cell type and whether *Apobec4* expression is regulated by miRNA-741. This could be analysed by cloning the *Apobec4* 3'UTR region upstream of a luciferase reporter gene and analysing the luciferase signals in the presence or absence of miRNA-741 expression. If the luciferase reporter experiments indicate miRNA-741 is regulating *Apobec4* then transgenic animals could be engineered lacking the target site for miRNA-741 to analyse the effects of *Apobec4* overexpression.

The polyclonal antibodies generated to recombinant A4 were able to recognise recombinant A4 produced with several different expression systems indicating the antibodies were specific for A4. Furthermore following affinity purification the

antibodies were no longer able to recognise the epitope tags that were present on the recombinant protein used to inoculate the guinea pigs indicating the affinity purification was successful. However endogenous A4 was not detected in testis or sperm extracts suggesting it is either expressed at a very low level or not expressed in these tissues. To determine the expression pattern of A4 protein transgenic animals could be engineered that express a reporter gene under the control of the A4 promoter. Alternatively, transgenic mice expressing a GFP-tagged A4 would also be informative as this would enable visualisation of the endogenous A4 protein. A GFP-tagged endogenous A4 would indicate whether the protein is expressed in mature sperm and whether A4 is present in the post-fertilisation embryo. Furthermore tagging the endogenous A4 would allow investigation of the subcellular localisation.

Full characterisation of the phenotype of the A4-null mice is ongoing, however they are healthy, viable and fertile indicating there are no obvious defects in spermatogenesis (Ian Adams, personal observation). Yet initial examination of A2- and A3-null mice showed no obvious phenotype (Mikl *et al.*, 2005). Whilst no function has been described for A2, a closer examination of the A2-null mice revealed a mild myopathic phenotype and diminished body mass indicating A2 has a biological function in maintaining muscle fibre ratios (Sato et al., 2009). A3 has been shown to have a crucial role in the immune system where it acts to inhibit retroviral infection and retroelement activation yet the A3-null mice have no obvious phenotype (Mikl et al., 2005) unless challenged (Okeoma et al., 2007). Therefore the phenotypic effect of loss of A4 may be subtle and require comprehensive analysis, and characterisation of the A4-null phenotype may indicate the biological function or substrate preference of A4.

The debate over the substrate specificity of AID highlights the difficulties faced when trying to characterise members of the AID/APOBEC family. AID has long been known to participate in the process of antibody diversification however clarification over whether AID deaminates DNA or RNA substrates has proved challenging. Many different approaches have been used to circumvent the need for

purified protein to analyse *in vitro*, including analysis of whether translation is required for CSR to occur in cell culture models (Begum et al., 2004). The substrate specificity of AID was eventually found to be ssDNA, although *in vitro* the enzyme was not active unless pre-treated with RNases to remove an inhibitory RNA (Bransteitter et al., 2003), leading to the hypothesis that RNA-binding may regulate the activity of AID on ssDNA via occupation of the catalytic active site. However, knowledge of the process that AID is involved in and analysis of the mutation spectrum at the immunoglobulin locus have helped clarify AID's role. In comparison, the lack of knowledge about the biological function of A4 has made identification of substrate specificity extremely challenging.

In contrast, characterisation of *Drosophila* ADAR has been aided greatly by the generation of *Adar*-null flies (Palladino et al., 2000b). Flies with the $Adar^{5G1}$ deletion show a range of phenotypes including locomotion defects and age-dependent neurodegeneration (Palladino et al., 2000b). Previously, substrates of *Drosophila* ADAR have been identified by the presence of guanosine in mRNAs that corresponds to adenosine residues in the DNA sequence, indicative of RNA editing. The discovery that the age-dependent neurodegeneration observed in *Adar*-null flies can be suppressed by expression of a catalytically inactive *Adar EA* transgene in the cholinergic neurons indicated that ADAR may have an additional deaminase-independent function in the CNS. However, identifying ADAR substrates that are bound by ADAR but not deaminated has proved challenging.

Previous analysis of $Adar^{5G1}$ flies indicated that the neurodegenerative phenotype observed was due to aberrant autophagy as no apoptotic cells were detected by TUNEL staining and large vacuoles were present in neuronal cell bodies (Leeanne McGurk, Thesis). However, microarray analysis of gene expression changes in 5 day old fly heads did not reveal changes in transcripts that are associated with either apoptosis or autophagy. Analysis of the GO terms associated with the altered genes indicated genes involved in the response to oxidative stress were altered indicating the onset of neuronal damage was occurring. Five day old flies were used to allow detection of gene expression changes associated with loss of *Adar* without secondary

changes induced by the neurodegenerative process. However, analysis of gene expression changes in older flies (10-15 days) is required to determine whether genes involved in programmed cell death are mis-regulated in *Adar*^{5G1} flies. This question was not addressed in the study described herein due to time and cost restrictions.

Using Affymetrix microarray technology to analyse gene expression levels I have identified a subset of transcripts which show altered expression in the absence of *Adar*, and are rescued in flies expressing either the active or the catalytically inactive *Adar* transgene. These results were confirmed by real time PCR; however it was unclear whether these transcript level changes were direct or indirect. I used two *in vitro* assays to characterise the interaction between ADAR and candidate transcripts further, however these assays indicated that ADAR binds to many transcripts *in vitro*. Previously *in vitro* analysis of ADAR RNA-binding has been performed on known substrates or perfect duplex dsRNA. This is the first demonstration that ADAR can bind to other pre-mRNAs with a similar affinity as known substrates. Furthermore promiscuous editing by ADAR was detected indicating these pre-mRNAs have a secondary structure that ADAR recognises, however no editing was detected in RNA isolated from wild type fly head. To analyse the interactions *in vivo* an RNA immunoprecipitation experiment was performed on transiently transfected *Drosophila* S2 cells. This indicated that ADAR does not bind to the candidate transcripts with the same affinity as a known ADAR substrate *Atx2*, however the experiment was difficult to optimise due to the low transfection efficiency of the S2 cell line. These results highlight the need for *in vivo* analysis of the interaction between ADAR and substrates and indicate that other factors affect the range of substrates ADAR interacts with *in vivo*. For example, competition for substrates with other RNA-binding proteins *in vivo* may restrict the set of transcripts which are bound by ADAR. However these results also suggest the altered gene expression observed in the *Adar*-null flies may occur through an indirect mechanism.

ADAR may affect the transcription of the altered genes through interactions with the RNA pol II CTD. However, analysis of the GO terms associated with the genes which show altered expression in *Adar*-null flies indicate these changes may be

induced in response to oxidative stress. The *Adar*-null flies do not show evidence of neurodegeneration at 5 days however ADAR may have a protective effect on neurons and when ADAR is absent the neurons may be more sensitive to oxidative stress. This is comparable to the role of ADAR2 in CA1 pyramidal neurons of the mouse brain, where ADAR2 mediates the response to ischemic insult by regulating cell death (Peng *et al.*, 2006). However in the mammalian system the regulation is mediated through editing of the GluR-B Q/R site. In the *Drosophila* system the role of ADAR in the stress response is unlikely to occur via site-specific editing of a single transcript as the neurodegenerative phenotype can be suppressed by expression of a catalytically inactive *Adar EA* transgene. Therefore the role of ADARs may be similar but the mechanism is different.

In further support of a role for ADAR in mediating the response to stress Gan et al have demonstrated that ADAR2 in mouse pancreatic islet cells is regulated in response to metabolic stress induced by a high fat diet (Gan *et al.*, 2006). Editing activity was also enhanced in mice fed on a high fat diet however it is not known what the role of ADAR2 in pancreatic islet cells is and why it is increased in times of metabolic stress. Further work revealed that ADAR2 is required for efficient exocytosis of secretory vesicles containing proteins such as insulin and chromogranin B (Yang *et al.*, 2010). An intact deaminase domain was required for efficient exocytosis demonstrating that RNA editing is required; however no substrates were identified. Decreased expression of Munc-18 and Synaptotagmin-7 which are key components of the exocytosis pathway was observed in ADAR2 deficient rat INS-1 cells; however it is unclear how ADAR2 affected their expression levels (Yang et al., 2010). This suggest an additional role for ADAR2 in regulation of vesicle transport and exocytosis, and whilst the homologs of *Munc*-18 and *Synaptotagmin*-7 were not altered according to the microarray data presented here, analysis of GO terms indicated expression of vesicle associated genes was altered in *Adar*-null flies.

Additionally the *Ilp3* gene showed decreased expression in *Adar*-null flies which could indicate a role for ADAR in the regulation of glucose metabolism. Further

work is required to distinguish the functional role of ADAR in regulating exocytosis and ADARs effect on regulation of the insulin pathway. An essential role for ADAR1 in mediating the interferon response in the developing liver has also been described recently (Hartner et al., 2009), such that ADAR1-null mice die at E12.5 due to interferon-induced apoptosis. Taken together these studies implicate ADARs in the response to cellular stress, although it is unclear what role ADAR plays and whether this effect is editing dependent or independent.

Numerous attempts have been made to characterise *Drosophila* ADAR substrates, however verification of novel substrates relies on identifying sequence changes indicative of RNA editing. We have proposed that *Drosophila* ADAR has an alternative editing-independent function; however it is unclear what this function is. To try and elucidate this editing-independent function of ADAR I have investigated changes in transcript levels and AS that occur in *Adar*-null flies. ADAR may have a much larger list of substrates than previously identified, however identification and verification of these substrates will require a new approach as these transcripts will not be edited. Cross-linked immunoprecipitation (CLIP) of transcripts bound by endogenous ADAR would provide a comprehensive list of substrates but may be technically challenging (Ule *et al.*, 2003). CLIP requires an antibody which is specific for *Drosophila* ADAR and this is currently being generated in the lab. An alternative would be to use the rescue lines which express an *Adar* transgene in the cholinergic neurons as these transgenes encode FLAG-tagged ADAR. However this may limit the list of substrates as endogenous *Adar* is expressed throughout the brain. In addition, CLIP requires a large amount of dissected brains as starting material, and the protocol for dissociation of cells and cross-linking would need to be optimised. However, following optimisation of the procedure CLIP should provide a comprehensive list of edited and unedited substrates of *Drosophila* ADAR. Once such a list is identified the editing-independent function of ADAR may become clearer. A similar approach has been used successfully to identify substrates of ADAR2 in mouse brain (Ohlson *et al.*, 2005), although cross-linking was not required here. This approach identified many known ADAR2 substrates, however

the authors do not discuss their false discovery rate and therefore it is unclear whether they identified transcripts which are bound by ADAR2 but are not edited.

It has been demonstrated that inosine-containing dsRNA can decrease gene expression and translation in *trans* (Scadden, 2007), however this is unlikely to account for the gene expression changes observed in *Adar*-null flies as transcript levels were both upregulated and downregulated. Furthermore the gene expression changes detected in *trans* caused by inosine-containing dsRNA were widespread and non-specific, occurring in both endogenous and reporter genes analysed. In contrast I have identified a significantly altered subset of mRNAs indicating the expression changes are specific. The loss of site-specific editing resulted in a subtle decrease in transcript levels for several known substrates of ADAR. This indicates that editing may stabilise these transcripts *in vivo*. However many genes showed stronger changes in expression than the known ADAR substrates indicating the loss of ADAR has far reaching effects.

Analysis of gene expression using an Affymetrix array identified a subset of transcripts which show altered transcript levels in *Adar*-null flies. However, AS array analysis identified a different subset of genes which were aberrantly spliced in the absence of ADAR. Known substrates of ADAR were not overrepresented in either group of transcripts indicating the loss of ADAR affects a wide range of transcripts. The results presented here are consistent with previous studies into the regulation of AS which have identified two distinct groups of genes, one group that is regulated at the transcript level and another group which is regulated via changes in AS (Ip et al., 2007). However it is unclear whether ADAR interacts directly with transcripts in either of these pathways. Technical difficulties with the RNA immunoprecipitation experiment prevented investigation into whether ADAR binds to the transcripts identified from the AS array analysis. Evidence in support of a direct interaction between ADAR and the transcripts which are aberrantly spliced in the absence of ADAR comes from the pattern of changes observed. Many of the aberrantly spliced genes showed alteration of single exon-exon junctions indicating these were not coordinated changes in known isoforms. Given the requirement of

dsRNA for ADAR binding it is interesting to hypothesise that dsRNA may be formed between the intron and the exon, similar to known editing site ECSs, and bound by ADAR in wild type flies. In the absence of ADAR these dsRNA structures may not be resolved efficiently and these structures may inhibit splicing. This has been shown to occur in the absence of the RNA helicase *maleless* where the *para* transcript undergoes a splicing catastrophe as the dsRNA structure required for editing is not resolved and this results in exclusion of the whole region from the mature transcript (Reenan *et al.*, 2000).

The observation that polyadenylation was altered in *Adar*-null flies was an unexpected result. These results could be due to miss-priming or spurious amplification products and require further validation; however the products were only detected in *Adar*-null flies implicating a role for ADAR in the generation of these products. Future experiments to characterise these changes include repeating the RACE experiment with polyA+ purified RNA to remove any partially processed transcripts. It would be interesting to analyse transcripts from these loci by Northern blot however the aberrant transcripts only accounted for a small fraction of the transcripts and may be difficult to detect on a Northern blot.

The alternative polyadenylation was not observed in rescued flies expressing an *Adar* transgene in the cholinergic neurons. Future experiments with RNA extracted from *Adar*-null flies expressing the catalytically inactive *Adar EA* transgene are required to determine whether the effect is dependent on RNA editing. Of note is that one of the aberrantly spliced transcripts (*CG13900*) encodes a protein component of the spliceosome (Mount and Salz, 2000) which contains an Interpro domain that is also found in the cleavage and polyadenylation stimulation factor (CPSF). Therefore the aberrant polyadenylation observed may arise from aberrant splicing of a component of the 3' end processing machinery. Further work is required to determine whether the *CG13900* protein level or activity is altered in *Adar*-null flies.

The array analysis presented in this thesis provides a valuable resource for identifying novel ADAR substrates in *Drosophila*. The AS array identifies the

specific region of the transcript which is affected by the loss of ADAR, however further bioinformatic analysis is required to determine whether the regions affected can form dsRNA required for ADAR to bind. A systems biology approach may be useful to integrate the expression array data and the AS array data to determine whether there are any common factors or pathways involved. This would allow integration of existing data on gene regulatory networks and protein interactions along with the array data described herein to provide a complete view of the biological impact of the loss of ADAR. A gene network is currently being compiled in the lab with the aim of superimposing data from both the expression array and the AS array to assess whether common regulatory networks of genes are affected by the absence of ADAR. This will provide a greater understanding of how ADAR interacts with other genes and may indicate which of the observed changes in transcript level and AS are direct and which may be indirect for future analysis. The GO term analysis described herein indicates these two groups of genes are functionally distinct; however it will be interesting to see whether the systems biology approach reveals that they are connected. The common feature shared between these groups may be a secondary structure RNA conformation which ADAR can bind to, but further bioinformatic analysis is required to analyse this.

Chapter 7: References

Agrawal, R., and Stormo, G.D. (2005). Editing efficiency of a Drosophila gene correlates with a distant splice site selection. Rna *11*, 563-566.

Al-Shahrour, F., Minguez, P., Tarraga, J., Montaner, D., Alloza, E., Vaquerizas, J.M., Conde, L., Blaschke, C., Vera, J., and Dopazo, J. (2006). BABELOMICS: a systems biology perspective in the functional annotation of genome-scale experiments. Nucleic Acids Res *34*, W472-476.

Alce, T.M., and Popik, W. (2004). APOBEC3G is incorporated into virus-like particles by a direct interaction with HIV-1 Gag nucleocapsid protein. J Biol Chem *279*, 34083-34086.

Anderson, P., and Kedersha, N. (2009). RNA granules: post-transcriptional and epigenetic modulators of gene expression. Nat Rev Mol Cell Biol *10*, 430-436.

Andrews, G.L., Tanglao, S., Farmer, W.T., Morin, S., Brotman, S., Berberoglu, M.A., Price, H., Fernandez, G.C., Mastick, G.S., Charron, F., and Kidd, T. (2008). Dscam guides embryonic axons by Netrin-dependent and -independent functions. Development *135*, 3839-3848.

Athanasiadis, A., Rich, A., and Maas, S. (2004). Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. PLoS Biol *2*, e391.

Backus, J.W., and Smith, H.C. (1991). Apolipoprotein B mRNA sequences 3' of the editing site are necessary and sufficient for editing and editosome assembly. Nucleic Acids Res *19*, 6781-6786.

Backus, J.W., and Smith, H.C. (1992). Three distinct RNA sequence elements are required for efficient apolipoprotein B (apoB) RNA editing in vitro. Nucleic Acids Res *20*, 6007-6014.

Bahadorani, S., Mukai, S., Egli, D., and Hilliker, A.J. (2008). Overexpression of metal-responsive transcription factor (MTF-1) in Drosophila melanogaster ameliorates life-span reductions associated with oxidative stress and metal toxicity. Neurobiol Aging.

Bass, B.L., and Weintraub, H. (1987). A developmentally regulated activity that unwinds RNA duplexes. Cell *48*, 607-613.

Bass, B.L., and Weintraub, H. (1988). An unwinding activity that covalently modifies its double-stranded RNA substrate. Cell *55*, 1089-1098.

Basto, R., Lau, J., Vinogradova, T., Gardiol, A., Woods, C.G., Khodjakov, A., and Raff, J.W. (2006). Flies without centrioles. Cell *125*, 1375-1386.

Beaujean, N., Taylor, J.E., McGarry, M., Gardner, J.O., Wilmut, I., Loi, P., Ptak, G., Galli, C., Lazzari, G., Bird, A., Young, L.E., and Meehan, R.R. (2004). The effect of interspecific oocytes on demethylation of sperm DNA. Proc Natl Acad Sci U S A *101*, 7636-7640.

Beghini, A., Ripamonti, C.B., Peterlongo, P., Roversi, G., Cairoli, R., Morra, E., and Larizza, L. (2000). RNA hyperediting and alternative splicing of hematopoietic cell phosphatase (PTPN6) gene in acute myeloid leukemia. Hum Mol Genet *9*, 2297-2304.

Begum, N.A., Kinoshita, K., Muramatsu, M., Nagaoka, H., Shinkura, R., and Honjo, T. (2004). De novo protein synthesis is required for activation-induced cytidine deaminase-dependent DNA cleavage in immunoglobulin class switch recombination. Proc Natl Acad Sci U S A *101*, 13003-13007.

Benjamini, Y., and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. The Annals of Statistics *29*, 1165-1188.

Benne, R., Van den Burg, J., Brakenhoff, J.P., Sloof, P., Van Boom, J.H., and Tromp, M.C. (1986). Major transcript of the frameshifted coxII gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. Cell *46*, 819-826.

Bennett, R.P., Diner, E., Sowden, M.P., Lees, J.A., Wedekind, J.E., and Smith, H.C. (2006). APOBEC-1 and AID are nucleo-cytoplasmic trafficking proteins but APOBEC3G cannot traffic. Biochem Biophys Res Commun *350*, 214-219.

Berek, C., and Milstein, C. (1988). The dynamic nature of the antibody repertoire. Immunol Rev *105*, 5-26.

Berget, S.M., Moore, C., and Sharp, P.A. (1977). Spliced segments at the 5' terminus of adenovirus 2 late mRNA. Proc Natl Acad Sci U S A *74*, 3171-3175.

Bernard, A., Ferhat, L., Dessi, F., Charton, G., Represa, A., Ben-Ari, Y., and Khrestchatisky, M. (1999). Q/R editing of the rat GluR5 and GluR6 kainate receptors in vivo and in vitro: evidence for independent developmental, pathological and cellular regulation. Eur J Neurosci *11*, 604-616.

Bernstein, E., Caudy, A.A., Hammond, S.M., and Hannon, G.J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature *409*, 363-366.

Betts, L., Xiang, S., Short, S.A., Wolfenden, R., and Carter, C.W., Jr. (1994). Cytidine deaminase. The 2.3 A crystal structure of an enzyme: transition-state analog complex. J Mol Biol *235*, 635-656.

Bhutani, N., Brady, J.J., Damian, M., Sacco, A., Corbel, S.Y., and Blau, H.M. (2009). Reprogramming towards pluripotency requires AID-dependent DNA demethylation. Nature.

Biniszkiewicz, D., Gribnau, J., Ramsahoye, B., Gaudet, F., Eggan, K., Humpherys, D., Mastrangelo, M.A., Jun, Z., Walter, J., and Jaenisch, R. (2002). Dnmt1 overexpression causes genomic hypermethylation, loss of imprinting, and embryonic lethality. Mol Cell Biol *22*, 2124-2135.

Black, D.L. (2003). Mechanisms of alternative pre-messenger RNA splicing. Annu Rev Biochem *72*, 291-336.

Blanc, V., Henderson, J.O., Kennedy, S., and Davidson, N.O. (2001). Mutagenesis of apobec-1 complementation factor reveals distinct domains that modulate RNA binding, protein-protein interaction with apobec-1, and complementation of C to U RNA-editing activity. J Biol Chem *276*, 46386-46393.

Blanc, V., Henderson, J.O., Newberry, E.P., Kennedy, S., Luo, J., and Davidson, N.O. (2005). Targeted deletion of the murine apobec-1 complementation factor (acf) gene results in embryonic lethality. Mol Cell Biol *25*, 7260-7269.

Blanchette, M., Green, R.E., MacArthur, S., Brooks, A.N., Brenner, S.E., Eisen, M.B., and Rio, D.C. (2009). Genome-wide analysis of alternative pre-mRNA splicing and RNA-binding specificities of the Drosophila hnRNP A/B family members. Mol Cell *33*, 438-449.

Blow, M., Futreal, P.A., Wooster, R., and Stratton, M.R. (2004). A survey of RNA editing in human brain. Genome Res *14*, 2379-2387.

Blow, M.J., Grocock, R.J., van Dongen, S., Enright, A.J., Dicks, E., Futreal, P.A., Wooster, R., and Stratton, M.R. (2006). RNA editing of human microRNAs. Genome Biol *7*, R27.

Bogerd, H.P., Doehle, B.P., Wiegand, H.L., and Cullen, B.R. (2004). A single amino acid difference in the host APOBEC3G protein controls the primate species specificity of HIV type 1 virion infectivity factor. Proc Natl Acad Sci U S A *101*, 3770-3774.

Bogerd, H.P., Wiegand, H.L., Doehle, B.P., Lueders, K.K., and Cullen, B.R. (2006a). APOBEC3A and APOBEC3B are potent inhibitors of LTR-retrotransposon function in human cells. Nucleic Acids Res *34*, 89-95.

Bogerd, H.P., Wiegand, H.L., Hulme, A.E., Garcia-Perez, J.L., O'Shea, K.S., Moran, J.V., and Cullen, B.R. (2006b). Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. Proc Natl Acad Sci U S A *103*, 8780-8785.

Borchert, G.M., Gilmore, B.L., Spengler, R.M., Xing, Y., Lanier, W., Bhattacharya, D., and Davidson, B.L. (2009). Adenosine deamination in human transcripts generates novel microRNA binding sites. Hum Mol Genet *18*, 4801-4807.

Boyd, J.B., and Harris, P.V. (1987). Isolation and characterization of a photorepair-deficient mutant in Drosophila melanogaster. Genetics *116*, 233-239.

Brand, A.H., and Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development *118*, 401-415.

Bransteitter, R., Pham, P., Scharff, M.D., and Goodman, M.F. (2003). Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. Proc Natl Acad Sci U S A *100*, 4102-4107.

Bransteitter, R., Prochnow, C., and Chen, X.S. (2009). The current structural and functional understanding of APOBEC deaminases. Cell Mol Life Sci *66*, 3137-3147.

Bratt, E., and Ohman, M. (2003). Coordination of editing and splicing of glutamate receptor pre-mRNA. Rna *9*, 309-318.

Breitling, R., Armengaud, P., Amtmann, A., and Herzyk, P. (2004). Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. FEBS Lett *573*, 83-92.

Brusa, R., Zimmermann, F., Koh, D.S., Feldmeyer, D., Gass, P., Seeburg, P.H., and Sprengel, R. (1995). Early-onset epilepsy and postnatal lethality associated with an editing-deficient GluR-B allele in mice. Science *270*, 1677-1680.

Buckingham, S.D., Kwak, S., Jones, A.K., Blackshaw, S.E., and Sattelle, D.B. (2008). Edited GluR2, a gatekeeper for motor neurone survival? Bioessays *30*, 1185-1192.

Burnashev, N., Monyer, H., Seeburg, P.H., and Sakmann, B. (1992). Divalent ion permeability of AMPA receptor channels is dominated by the edited form of a single subunit. Neuron *8*, 189-198.

Burns, C.M., Chu, H., Rueter, S.M., Hutchinson, L.K., Canton, H., Sanders-Bush, E., and Emeson, R.B. (1997). Regulation of serotonin-2C receptor G-protein coupling by RNA editing. Nature *387*, 303-308.

Caceres, J.F., and Kornblihtt, A.R. (2002). Alternative splicing: multiple control mechanisms and involvement in human disease. Trends Genet *18*, 186-193.

Carthew, R.W., and Sontheimer, E.J. (2009). Origins and Mechanisms of miRNAs and siRNAs. Cell *136*, 642-655.

Casey, J.L., and Gerin, J.L. (1995). Hepatitis D virus RNA editing: specific modification of adenosine in the antigenomic RNA. J Virol *69*, 7593-7600.

Cenci, C., Barzotti, R., Galeano, F., Corbelli, S., Rota, R., Massimi, L., Di Rocco, C., O'Connell, M.A., and Gallo, A. (2008). Down-regulation of RNA editing in pediatric astrocytomas: ADAR2 editing activity inhibits cell migration and proliferation. J Biol Chem *283*, 7251-7260.

Chalmel, F., Rolland, A.D., Niederhauser-Wiederkehr, C., Chung, S.S., Demougin, P., Gattiker, A., Moore, J., Patard, J.J., Wolgemuth, D.J., Jegou, B., and Primig, M. (2007). The conserved transcriptome in human and rodent male gametogenesis. Proc Natl Acad Sci U S A *104*, 8346-8351.

Champagne, M.B., Edwards, K.A., Erickson, H.P., and Kiehart, D.P. (2000). Drosophila stretchin-MLCK is a novel member of the Titin/Myosin light chain kinase family. J Mol Biol *300*, 759-777.

Chang, Y., Kong, Q., Shan, X., Tian, G., Ilieva, H., Cleveland, D.W., Rothstein, J.D., Borchelt, D.R., Wong, P.C., and Lin, C.L. (2008). Messenger RNA oxidation occurs early in disease pathogenesis and promotes motor neuron degeneration in ALS. PLoS One *3*, e2849.

Chen, C., Jack, J., and Garofalo, R.S. (1996). The Drosophila insulin receptor is required for normal growth. Endocrinology *137*, 846-856.

Chen, C.X., Cho, D.S., Wang, Q., Lai, F., Carter, K.C., and Nishikura, K. (2000). A third member of the RNA-specific adenosine deaminase gene family, ADAR3, contains both single- and double-stranded RNA binding domains. Rna *6*, 755-767.

Chen, J., Lakshmi, G.G., Hays, D.L., McDowell, K.M., Ma, E., and Vaughn, J.C. (2009). Spatial and temporal expression of dADAR mRNA and protein isoforms during embryogenesis in Drosophila melanogaster. Differentiation *78*, 312-320.

Chen, J.M., Stenson, P.D., Cooper, D.N., and Ferec, C. (2005). A systematic analysis of LINE-1 endonuclease-dependent retrotranspositional events causing human genetic disease. Hum Genet *117*, 411-427.

Chen, L., Rio, D.C., Haddad, G.G., and Ma, E. (2004). Regulatory role of dADAR in ROS metabolism in Drosophila CNS. Brain Res Mol Brain Res *131*, 93-100.

Cherbas, P., Cherbas, L., and Williams, C.M. (1977). Induction of acetylcholinesterase activity by beta-ecdysone in a Drosophila cell line. Science *197*, 275-277.

Chesebro, B., and Wehrly, K. (1979). Identification of a non-H-2 gene (Rfv-3) influencing recovery from viremia and leukemia induced by Friend virus complex. Proc Natl Acad Sci U S A *76*, 425-429.

Chester, A., Somasekaram, A., Tzimina, M., Jarmuz, A., Gisbourne, J., O'Keefe, R., Scott, J., and Navaratnam, N. (2003). The apolipoprotein B mRNA editing complex performs a multifunctional cycle and suppresses nonsense-mediated decay. Embo J *22*, 3971-3982.

Chester, A., Weinreb, V., Carter, C.W., Jr., and Navaratnam, N. (2004). Optimization of apolipoprotein B mRNA editing by APOBEC1 apoenzyme and the role of its auxiliary factor, ACF. Rna *10*, 1399-1411.

Chilibeck, K.A., Wu, T., Liang, C., Schellenberg, M.J., Gesner, E.M., Lynch, J.M., and MacMillan, A.M. (2006). FRET analysis of in vivo dimerization by RNA-editing enzymes. J Biol Chem *281*, 16530-16535.

Chiu, Y.L., Soros, V.B., Kreisberg, J.F., Stopak, K., Yonemoto, W., and Greene, W.C. (2005). Cellular APOBEC3G restricts HIV-1 infection in resting CD4+ T cells. Nature *435*, 108-114.

Chiu, Y.L., Witkowska, H.E., Hall, S.C., Santiago, M., Soros, V.B., Esnault, C., Heidmann, T., and Greene, W.C. (2006). High-molecular-mass APOBEC3G complexes restrict Alu retrotransposition. Proc Natl Acad Sci U S A *103*, 15588-15593.

Cho, D.S., Yang, W., Lee, J.T., Shiekhattar, R., Murray, J.M., and Nishikura, K. (2003). Requirement of dimerization for RNA editing activity of adenosine deaminases acting on RNA. J Biol Chem *278*, 17093-17102.

Chow, L.T., Gelinas, R.E., Broker, T.R., and Roberts, R.J. (1977). An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. Cell *12*, 1-8.

Clutterbuck, D.R., Leroy, A., O'Connell, M.A., and Semple, C.A. (2005). A bioinformatic screen for novel A-I RNA editing sites reveals recoding editing in BC10. Bioinformatics *21*, 2590-2595.

Coker, H.A., Morgan, H.D., and Petersen-Mahrt, S.K. (2006). Genetic and in vitro assays of DNA deamination. Methods Enzymol *408*, 156-170.

Connolly, C.M., Dearth, A.T., and Braun, R.E. (2005). Disruption of murine Tenr results in teratospermia and male infertility. Dev Biol *278*, 13-21.

Conticello, S.G., Thomas, C.J., Petersen-Mahrt, S.K., and Neuberger, M.S. (2005). Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. Mol Biol Evol *22*, 367-377.

Cordaux, R., and Batzer, M.A. (2009). The impact of retrotransposons on human genome evolution. Nat Rev Genet *10*, 691-703.

Costinean, S., Zanesi, N., Pekarsky, Y., Tili, E., Volinia, S., Heerema, N., and Croce, C.M. (2006). Pre-B cell proliferation and lymphoblastic leukemia/high-grade lymphoma in E(mu)-miR155 transgenic mice. Proc Natl Acad Sci U S A *103*, 7024-7029.

D'Souza, I., Poorkaj, P., Hong, M., Nochlin, D., Lee, V.M., Bird, T.D., and Schellenberg, G.D. (1999). Missense and silent tau gene mutations cause frontotemporal dementia with parkinsonism-chromosome 17 type, by affecting multiple alternative RNA splicing regulatory elements. Proc Natl Acad Sci U S A *96*, 5598-5603.

Dance, G.S., Beemiller, P., Yang, Y., Mater, D.V., Mian, I.S., and Smith, H.C. (2001). Identification of the yeast cytidine deaminase CDD1 as an orphan C-->U RNA editase. Nucleic Acids Res *29*, 1772-1780.

Daniel, C., and Ohman, M. (2009). RNA editing and its impact on GABAA receptor function. Biochem Soc Trans *37*, 1399-1403.

Davidson, N.O., and Shelness, G.S. (2000). APOLIPOPROTEIN B: mRNA editing, lipoprotein assembly, and presecretory degradation. Annu Rev Nutr *20*, 169-193.

Dawson, T.R., Sansam, C.L., and Emeson, R.B. (2004). Structure and sequence determinants required for the RNA editing of ADAR2 substrates. J Biol Chem *279*, 4941-4951.

de Rooij, D.G. (1998). Stem cells in the testis. Int J Exp Pathol *79*, 67-80.

Desterro, J.M., Keegan, L.P., Jaffray, E., Hay, R.T., O'Connell, M.A., and Carmo-Fonseca, M. (2005). SUMO-1 modification alters ADAR1 editing activity. Mol Biol Cell *16*, 5115-5126.

Desterro, J.M., Keegan, L.P., Lafarga, M., Berciano, M.T., O'Connell, M., and Carmo-Fonseca, M. (2003). Dynamic association of RNA-editing enzymes with the nucleolus. J Cell Sci *116*, 1805-1818.

Doehle, B.P., Schafer, A., Wiegand, H.L., Bogerd, H.P., and Cullen, B.R. (2005). Differential sensitivity of murine leukemia virus to APOBEC3-mediated inhibition is governed by virion exclusion. J Virol *79*, 8201-8207.

Dorn, R., and Krauss, V. (2003). The modifier of mdg4 locus in Drosophila: functional complexity is resolved by trans splicing. Genetica *117*, 165-177.

Dorsett, Y., McBride, K.M., Jankovic, M., Gazumyan, A., Thai, T.H., Robbiani, D.F., Di Virgilio, M., San-Martin, B.R., Heidkamp, G., Schwickert, T.A., Eisenreich, T., Rajewsky, K., and Nussenzweig, M.C. (2008). MicroRNA-155 suppresses activation-induced cytidine deaminase-mediated Myc-Igh translocation. Immunity *28*, 630-638.

Echalier, G., and Ohanessian, A. (1969). [Isolation, in tissue culture, of Drosophila melangaster cell lines]. C R Acad Sci Hebd Seances Acad Sci D *268*, 1771-1773.

Eis, P.S., Tam, W., Sun, L., Chadburn, A., Li, Z., Gomez, M.F., Lund, E., and Dahlberg, J.E. (2005). Accumulation of miR-155 and BIC RNA in human B cell lymphomas. Proc Natl Acad Sci U S A *102*, 3627-3632.

Esnault, C., Heidmann, O., Delebecque, F., Dewannieux, M., Ribet, D., Hance, A.J., Heidmann, T., and Schwartz, O. (2005). APOBEC3G cytidine deaminase inhibits retrotransposition of endogenous retroviruses. Nature *433*, 430-433.

Farese, R.V., Jr., Veniant, M.M., Cham, C.M., Flynn, L.M., Pierotti, V., Loring, J.F., Traber, M., Ruland, S., Stokowski, R.S., Huszar, D., and Young, S.G. (1996). Phenotypic analysis of mice expressing exclusively apolipoprotein B48 or apolipoprotein B100. Proc Natl Acad Sci U S A *93*, 6393-6398.

Feng, Y., Sansam, C.L., Singh, M., and Emeson, R.B. (2006). Altered RNA editing in mice lacking ADAR2 autoregulation. Mol Cell Biol *26*, 480-488.

ffrench-Constant, R.H., and Rocheleau, T.A. (1993). Drosophila gamma-aminobutyric acid receptor gene Rdl shows extensive alternative splicing. J Neurochem *60*, 2323-2326.

ffrench-Constant, R.H., Steichen, J.C., Rocheleau, T.A., Aronstein, K., and Roush, R.T. (1993). A single-amino acid substitution in a gamma-aminobutyric acid subtype A receptor locus is associated with cyclodiene insecticide resistance in Drosophila populations. Proc Natl Acad Sci U S A *90*, 1957-1961.

Fierro-Monti, I., and Mathews, M.B. (2000). Proteins binding to duplexed RNA: one motif, multiple functions. Trends Biochem Sci *25*, 241-246.

Flomen, R., Knight, J., Sham, P., Kerwin, R., and Makoff, A. (2004). Evidence that RNA editing modulates splice site selection in the 5-HT2C receptor gene. Nucleic Acids Res *32*, 2113-2122.

Friedman, R.C., Farh, K.K., Burge, C.B., and Bartel, D.P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. Genome Res *19*, 92-105.

Gallo, A., Keegan, L.P., Ring, G.M., and O'Connell, M.A. (2003). An ADAR that edits transcripts encoding ion channel subunits functions as a dimer. Embo J *22*, 3421-3430.

Gallois-Montbrun, S., Kramer, B., Swanson, C.M., Byers, H., Lynham, S., Ward, M., and Malim, M.H. (2007). Antiviral protein APOBEC3G localizes to ribonucleoprotein complexes found in P bodies and stress granules. J Virol *81*, 2165-2178.

Gan, Z., Zhao, L., Yang, L., Huang, P., Zhao, F., Li, W., and Liu, Y. (2006). RNA editing by ADAR2 is metabolically regulated in pancreatic islets and beta-cells. J Biol Chem *281*, 33386-33394.

Gawande, B., Robida, M.D., Rahn, A., and Singh, R. (2006). Drosophila Sex-lethal protein mediates polyadenylation switching in the female germline. Embo J *25*, 1263-1272.

Geiss, G., Jin, G., Guo, J., Bumgarner, R., Katze, M.G., and Sen, G.C. (2001). A comprehensive view of regulation of gene expression by double-stranded RNA-mediated cell signaling. J Biol Chem *276*, 30178-30182.

Gerber, A., Grosjean, H., Melcher, T., and Keller, W. (1998). Tad1p, a yeast tRNA-specific adenosine deaminase, is related to the mammalian pre-mRNA editing enzymes ADAR1 and ADAR2. Embo J *17*, 4780-4789.

Gerber, A., O'Connell, M.A., and Keller, W. (1997). Two forms of human double-stranded RNA-specific editase 1 (hRED1) generated by the insertion of an Alu cassette. Rna *3*, 453-463.

Gerber, A.P., and Keller, W. (1999). An adenosine deaminase that generates inosine at the wobble position of tRNAs. Science *286*, 1146-1149.

Gerber, A.P., and Keller, W. (2001). RNA editing by base deamination: more enzymes, more targets, new mysteries. Trends Biochem Sci *26*, 376-384.

Girardot, F., Monnier, V., and Tricoire, H. (2004). Genome wide analysis of common and specific stress responses in adult drosophila melanogaster. BMC Genomics *5*, 74.

Gonzalez, M.C., Suspene, R., Henry, M., Guetard, D., Wain-Hobson, S., and Vartanian, J.P. (2009). Human APOBEC1 cytidine deaminase edits HBV DNA. Retrovirology *6*, 96.

Grauso, M., Reenan, R.A., Culetto, E., and Sattelle, D.B. (2002). Novel putative nicotinic acetylcholine receptor subunit genes, Dalpha5, Dalpha6 and Dalpha7, in Drosophila melanogaster identify a new and highly conserved target of adenosine deaminase acting on RNA-mediated A-to-I pre-mRNA editing. Genetics *160*, 1519-1533.

Graveley, B.R. (2005). Mutually exclusive splicing of the insect Dscam pre-mRNA directed by competing intronic RNA secondary structures. Cell *123*, 65-73.

Greger, I.H., Akamine, P., Khatri, L., and Ziff, E.B. (2006). Developmentally regulated, combinatorial RNA processing modulates AMPA receptor biogenesis. Neuron *51*, 85-97.

Greger, I.H., Khatri, L., Kong, X., and Ziff, E.B. (2003). AMPA receptor tetramerization is mediated by Q/R editing. Neuron *40*, 763-774.

Greger, I.H., Khatri, L., and Ziff, E.B. (2002). RNA editing at arg607 controls AMPA receptor exit from the endoplasmic reticulum. Neuron *34*, 759-772.

Gregory, R.I., Chendrimada, T.P., Cooch, N., and Shiekhattar, R. (2005). Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. Cell *123*, 631-640.

Gronke, S., Clarke, D.F., Broughton, S., Andrews, T.D., and Partridge, L. (2010). Molecular evolution and functional characterization of Drosophila insulin-like peptides. PLoS Genet *6*, e1000857.

Gruber, A.R., Lorenz, R., Bernhart, S.H., Neubock, R., and Hofacker, I.L. (2008). The Vienna RNA websuite. Nucleic Acids Res *36*, W70-74.

Gu, R., Zhang, Z., DeCerbo, J.N., and Carmichael, G.G. (2009). Gene regulation by sense-antisense overlap of polyadenylation signals. Rna *15*, 1154-1163.

Gurney, M.E., Pu, H., Chiu, A.Y., Dal Canto, M.C., Polchow, C.Y., Alexander, D.D., Caliendo, J., Hentati, A., Kwon, Y.W., Deng, H.X., and et al. (1994). Motor neuron degeneration in mice that express a human Cu,Zn superoxide dismutase mutation. Science *264*, 1772-1775.

Hajkova, P., Erhardt, S., Lane, N., Haaf, T., El-Maarri, O., Reik, W., Walter, J., and Surani, M.A. (2002). Epigenetic reprogramming in mouse primordial germ cells. Mech Dev *117*, 15-23.

Hallegger, M., Taschner, A., and Jantsch, M.F. (2006). RNA aptamers binding the double-stranded RNA-binding domain. Rna *12*, 1993-2004.

Hanrahan, C.J., Palladino, M.J., Ganetzky, B., and Reenan, R.A. (2000). RNA editing of the Drosophila para Na(+) channel transcript. Evolutionary conservation and developmental regulation. Genetics *155*, 1149-1160.

Harris, R.S., Petersen-Mahrt, S.K., and Neuberger, M.S. (2002). RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. Mol Cell *10*, 1247-1253.

Hartmann, B., Castelo, R., Blanchette, M., Boue, S., Rio, D.C., and Valcarcel, J. (2009). Global analysis of alternative splicing regulation by insulin and wingless signaling in Drosophila cells. Genome Biol *10*, R11.

Hartner, J.C., Schmittwolf, C., Kispert, A., Muller, A.M., Higuchi, M., and Seeburg, P.H. (2004). Liver disintegration in the mouse embryo caused by deficiency in the RNA-editing enzyme ADAR1. J Biol Chem *279*, 4894-4902.

Hartner, J.C., Walkley, C.R., Lu, J., and Orkin, S.H. (2009). ADAR1 is essential for the maintenance of hematopoiesis and suppression of interferon signaling. Nat Immunol *10*, 109-115.

Hattori, D., Chen, Y., Matthews, B.J., Salwinski, L., Sabatti, C., Grueber, W.B., and Zipursky, S.L. (2009). Robust discrimination between self and non-self neurites requires thousands of Dscam1 isoforms. Nature *461*, 644-648.

Haynes, S.R., ed. (1999). RNA Interaction Protocols (Humana Press Inc).

Heale, B.S., Keegan, L.P., McGurk, L., Michlewski, G., Brindle, J., Stanton, C.M., Caceres, J.F., and O'Connell, M.A. (2009). Editing independent effects of ADARs on the miRNA/siRNA pathways. Embo J *28*, 3145-3156.

Heisenberg, M. (2003). Mushroom body memoir: from maps to models. Nat Rev Neurosci *4*, 266-275.

Herbert, A., and Rich, A. (2001). The role of binding domains for dsRNA and Z-DNA in the in vivo editing of minimal substrates by ADAR1. Proc Natl Acad Sci U S A *98*, 12132-12137.

Higuchi, M., Maas, S., Single, F.N., Hartner, J., Rozov, A., Burnashev, N., Feldmeyer, D., Sprengel, R., and Seeburg, P.H. (2000). Point mutation in an AMPA receptor gene rescues lethality in mice deficient in the RNA-editing enzyme ADAR2. Nature *406*, 78-81.

Higuchi, M., Single, F.N., Kohler, M., Sommer, B., Sprengel, R., and Seeburg, P.H. (1993). RNA editing of AMPA receptor subunit GluR-B: a base-paired intron-exon structure determines position and efficiency. Cell *75*, 1361-1370.

Hirano, K., Young, S.G., Farese, R.V., Jr., Ng, J., Sande, E., Warburton, C., Powell-Braxton, L.M., and Davidson, N.O. (1996). Targeted disruption of the mouse apobec-1 gene abolishes apolipoprotein B mRNA editing and eliminates apolipoprotein B48. J Biol Chem *271*, 9887-9890.

Hoffman, B.E., and Lis, J.T. (2000). Pre-mRNA splicing by the essential Drosophila protein B52: tissue and target specificity. Mol Cell Biol *20*, 181-186.

Hoopengardner, B., Bhalla, T., Staber, C., and Reenan, R. (2003). Nervous system targets of RNA editing identified by comparative genomics. Science *301*, 832-836.

Hough, R.F., and Bass, B.L. (1997). Analysis of Xenopus dsRNA adenosine deaminase cDNAs reveals similarities to DNA methyltransferases. Rna *3*, 356-370.

Hundley, H.A., Krauchuk, A.A., and Bass, B.L. (2008). C. elegans and H. sapiens mRNAs with edited 3' UTRs are present on polysomes. Rna *14*, 2050-2060.

Huthoff, H., and Malim, M.H. (2005). Cytidine deamination and resistance to retroviral infection: towards a structural understanding of the APOBEC proteins. Virology *334*, 147-153.

Ikeya, T., Galic, M., Belawat, P., Nairz, K., and Hafen, E. (2002). Nutrient-dependent expression of insulin-like peptides from neuroendocrine cells in the CNS contributes to growth regulation in Drosophila. Curr Biol *12*, 1293-1300.

Ip, J.Y., Tong, A., Pan, Q., Topp, J.D., Blencowe, B.J., and Lynch, K.W. (2007). Global analysis of alternative splicing during T-cell activation. Rna *13*, 563-572.

Jaikaran, D.C., Collins, C.H., and MacMillan, A.M. (2002). Adenosine to inosine editing by ADAR2 requires formation of a ternary complex on the GluR-B R/G site. J Biol Chem *277*, 37624-37629.

Jarmuz, A., Chester, A., Bayliss, J., Gisbourne, J., Dunham, I., Scott, J., and Navaratnam, N. (2002). An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22. Genomics *79*, 285-296.

Jepson, J.E., and Reenan, R.A. (2009). Adenosine-to-inosine genetic recoding is required in the adult stage nervous system for coordinated behavior in Drosophila. J Biol Chem *284*, 31391-31400.

Jia, Z., Agopyan, N., Miu, P., Xiong, Z., Henderson, J., Gerlai, R., Taverna, F.A., Velumian, A., MacDonald, J., Carlen, P., Abramow-Newerly, W., and Roder, J. (1996). Enhanced LTP in mice deficient in the AMPA receptor GluR2. Neuron *17*, 945-956.

Jin, Y., Tian, N., Cao, J., Liang, J., Yang, Z., and Lv, J. (2007). RNA editing and alternative splicing of the insect nAChR subunit alpha6 transcript: evolutionary conservation, divergence and regulation. BMC Evol Biol *7*, 98.

Jones, A.K., Buckingham, S.D., Papadaki, M., Yokota, M., Sattelle, B.M., Matsuda, K., and Sattelle, D.B. (2009). Splice-variant- and stage-specific RNA editing of the Drosophila GABA receptor modulates agonist potency. J Neurosci *29*, 4287-4292.

Kallman, A.M., Sahlin, M., and Ohman, M. (2003). ADAR2 A-->I editing: site selectivity and editing efficiency are separate events. Nucleic Acids Res *31*, 4874-4881.

Kao, S., Khan, M.A., Miyagi, E., Plishka, R., Buckler-White, A., and Strebel, K. (2003). The human immunodeficiency virus type 1 Vif protein reduces intracellular expression and inhibits packaging of APOBEC3G (CEM15), a cellular inhibitor of virus infectivity. J Virol *77*, 11398-11407.

Kask, K., Zamanillo, D., Rozov, A., Burnashev, N., Sprengel, R., and Seeburg, P.H. (1998). The AMPA receptor subunit GluR-B in its Q/R site-unedited form is not essential for brain development and function. Proc Natl Acad Sci U S A *95*, 13777-13782.

Kato, Y., Kaneda, M., Hata, K., Kumaki, K., Hisano, M., Kohara, Y., Okano, M., Li, E., Nozaki, M., and Sasaki, H. (2007). Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse. Hum Mol Genet *16*, 2272-2280.

Kawahara, Y., Ito, K., Sun, H., Aizawa, H., Kanazawa, I., and Kwak, S. (2004). Glutamate receptors: RNA editing and death of motor neurons. Nature *427*, 801.

Kawahara, Y., Megraw, M., Kreider, E., Iizasa, H., Valente, L., Hatzigeorgiou, A.G., and Nishikura, K. (2008). Frequency and fate of microRNA editing in human brain. Nucleic Acids Res *36*, 5270-5280.

Kawahara, Y., Zinshteyn, B., Chendrimada, T.P., Shiekhattar, R., and Nishikura, K. (2007a). RNA editing of the microRNA-151 precursor blocks cleavage by the Dicer-TRBP complex. EMBO Rep *8*, 763-769.

Kawahara, Y., Zinshteyn, B., Sethupathy, P., Iizasa, H., Hatzigeorgiou, A.G., and Nishikura, K. (2007b). Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. Science *315*, 1137-1140.

Kawakubo, K., and Samuel, C.E. (2000). Human RNA-specific adenosine deaminase (ADAR1) gene specifies transcripts that initiate from a constitutively active alternative promoter. Gene *258*, 165-172.

Keegan, L.P., Brindle, J., Gallo, A., Leroy, A., Reenan, R.A., and O'Connell, M.A. (2005). Tuning of RNA editing by ADAR is required in Drosophila. Embo J *24*, 2183-2193.

Keegan, L.P., Gerber, A.P., Brindle, J., Leemans, R., Gallo, A., Keller, W., and O'Connell, M.A. (2000). The properties of a tRNA-specific adenosine deaminase from Drosophila melanogaster support an evolutionary link between pre-mRNA editing and tRNA modification. Mol Cell Biol *20*, 825-833.

Keegan, L.P., Leroy, A., Sproul, D., and O'Connell, M.A. (2004). Adenosine deaminases acting on RNA (ADARs): RNA-editing enzymes. Genome Biol *5*, 209.

Kelleher, J.E., and Raleigh, E.A. (1991). A novel activity in Escherichia coli K-12 that directs restriction of DNA modified at CG dinucleotides. J Bacteriol *173*, 5220-5223.

Kim, D.D., Kim, T.T., Walsh, T., Kobayashi, Y., Matise, T.C., Buyske, S., and Gabriel, A. (2004). Widespread RNA editing of embedded alu elements in the human transcriptome. Genome Res *14*, 1719-1725.

Kim, V.N. (2005). MicroRNA biogenesis: coordinated cropping and dicing. Nat Rev Mol Cell Biol *6*, 376-385.

Kinomoto, M., Kanno, T., Shimura, M., Ishizaka, Y., Kojima, A., Kurata, T., Sata, T., and Tokunaga, K. (2007). All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition. Nucleic Acids Res *35*, 2955-2964.

Klaue, Y., Kallman, A.M., Bonin, M., Nellen, W., and Ohman, M. (2003). Biochemical analysis and scanning force microscopy reveal productive and nonproductive ADAR2 binding to RNA substrates. Rna *9*, 839-846.

Klose, R.J., and Bird, A.P. (2006). Genomic DNA methylation: the mark and its mediators. Trends Biochem Sci *31*, 89-97.

Knight, S.W., and Bass, B.L. (2002). The role of RNA editing by ADARs in RNAi. Mol Cell *10*, 809-817.

Kobayashi, M., Takaori-Kondo, A., Miyauchi, Y., Iwai, K., and Uchiyama, T. (2005). Ubiquitination of APOBEC3G by an HIV-1 Vif-Cullin5-Elongin B-Elongin C complex is essential for Vif function. J Biol Chem *280*, 18573-18578.

Kock, J., and Blum, H.E. (2008). Hypermutation of hepatitis B virus genomes by APOBEC3G, APOBEC3C and APOBEC3H. J Gen Virol *89*, 1184-1191.

Koike, M., Tsukada, S., Tsuzuki, K., Kijima, H., and Ozawa, S. (2000). Regulation of kinetic properties of GluR2 AMPA receptor channels by alternative splicing. J Neurosci *20*, 2166-2174.

Kondo, T., Suzuki, T., Ito, S., Kono, M., Negoro, T., and Tomita, Y. (2008). Dyschromatosis symmetrica hereditaria associated with neurological disorders. J Dermatol *35*, 662-666.

Kozak, S.L., Marin, M., Rose, K.M., Bystrom, C., and Kabat, D. (2006). The anti-HIV-1 editing enzyme APOBEC3G binds HIV-1 RNA and messenger RNAs that shuttle between polysomes and stress granules. J Biol Chem *281*, 29105-29119.

Krawczak, M., Thomas, N.S., Hundrieser, B., Mort, M., Wittig, M., Hampe, J., and Cooper, D.N. (2007). Single base-pair substitutions in exon-intron junctions of human genes: nature, distribution, and consequences for mRNA splicing. Hum Mutat *28*, 150-158.

Kretzschmar, D., Hasan, G., Sharma, S., Heisenberg, M., and Benzer, S. (1997). The swiss cheese mutant causes glial hyperwrapping and brain degeneration in Drosophila. J Neurosci *17*, 7425-7432.

Kuner, R., Groom, A.J., Bresink, I., Kornau, H.C., Stefovska, V., Muller, G., Hartmann, B., Tschauner, K., Waibel, S., Ludolph, A.C., Ikonomidou, C., Seeburg, P.H., and Turski, L. (2005). Late-onset motoneuron disease caused by a functionally modified AMPA receptor subunit. Proc Natl Acad Sci U S A *102*, 5826-5831.

Kwak, S., and Kawahara, Y. (2005). Deficient RNA editing of GluR2 and neuronal death in amyotropic lateral sclerosis. J Mol Med *83*, 110-120.

Kwak, S., and Weiss, J.H. (2006). Calcium-permeable AMPA channels in neurodegenerative disease and ischemia. Curr Opin Neurobiol *16*, 281-287.

Labarga, A., Valentin, F., Anderson, M., and Lopez, R. (2007). Web services at the European bioinformatics institute. Nucleic Acids Res *35*, W6-11.

Lai, F., Chen, C.X., Carter, K.C., and Nishikura, K. (1997). Editing of glutamate receptor B subunit ion channel RNAs by four alternatively spliced DRADA2 double-stranded RNA adenosine deaminases. Mol Cell Biol *17*, 2413-2424.

Lai, F., Drakas, R., and Nishikura, K. (1995). Mutagenic analysis of double-stranded RNA adenosine deaminase, a candidate enzyme for RNA editing of glutamate-gated ion channel transcripts. J Biol Chem *270*, 17098-17105.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-

Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J.C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra, M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissoe, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R.A., Muzny, D.M., Scherer, S.E., Bouck, J.B., Sodergren, E.J., Worley, K.C., Rives, C.M., Gorrell, J.H., Metzker, M.L., Naylor, S.L., Kucherlapati, R.S., Nelson, D.L., Weinstock, G.M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D.R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H.M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R.W., Federspiel, N.A., Abola, A.P., Proctor, M.J., Myers, R.M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D.R., Olson, M.V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G.A., Athanasiou, M., Schultz, R., Roe, B.A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W.R., de la Bastide, M., Dedhia, N., Blocker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J.A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D.G., Burge, C.B., Cerutti, L., Chen, H.C., Church, D., Clamp, M., Copley, R.R., Doerks, T., Eddy, S.R., Eichler, E.E., Furey, T.S., Galagan, J., Gilbert, J.G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L.S., Jones, T.A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W.J., Kitts, P., Koonin, E.V., Korf, I., Kulp, D., Lancet, D., Lowe, T.M., McLysaght, A., Mikkelsen, T., Moran, J.V., Mulder, N., Pollara, V.J., Ponting, C.P., Schuler, G., Schultz, J., Slater, G., Smit, A.F., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y.I., Wolfe, K.H., Yang, S.P., Yeh, R.F., Collins, F., Guyer, M.S., Peterson, J., Felsenfeld, A., Wetterstrand, K.A., Patrinos, A., Morgan, M.J., de Jong, P., Catanese, J.J., Osoegawa, K., Shizuya, H., Choi, S., and Chen, Y.J. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Lane, N., Dean, W., Erhardt, S., Hajkova, P., Surani, A., Walter, J., and Reik, W. (2003). Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. Genesis *35*, 88-93.

Lau, P.P., Xiong, W.J., Zhu, H.J., Chen, S.H., and Chan, L. (1991). Apolipoprotein B mRNA editing is an intranuclear event that occurs posttranscriptionally coincident with splicing and polyadenylation. J Biol Chem *266*, 20550-20554.

Lau, P.P., Zhu, H.J., Baldini, A., Charnsangavej, C., and Chan, L. (1994). Dimeric structure of a human apolipoprotein B mRNA editing protein and cloning and chromosomal localization of its gene. Proc Natl Acad Sci U S A *91*, 8522-8526.

Laurencikiene, J., Kallman, A.M., Fong, N., Bentley, D.L., and Ohman, M. (2006). RNA editing and alternative splicing: the importance of co-transcriptional coordination. EMBO Rep *7*, 303-307.

Lawson, K.A., and Pedersen, R.A. (1992). Clonal analysis of cell fate during gastrulation and early neurulation in the mouse. Ciba Found Symp *165*, 3-21; discussion 21-26.

Lebecque, S.G., and Gearhart, P.J. (1990). Boundaries of somatic mutation in rearranged immunoglobulin genes: 5' boundary is near the promoter, and 3' boundary is approximately 1 kb from V(D)J gene. J Exp Med *172*, 1717-1727.

Lee, Y.S., Nakahara, K., Pham, J.W., Kim, K., He, Z., Sontheimer, E.J., and Carthew, R.W. (2004). Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. Cell *117*, 69-81.

Lehmann, K.A., and Bass, B.L. (2000). Double-stranded RNA adenosine deaminases ADAR1 and ADAR2 have overlapping specificities. Biochemistry *39*, 12875-12884.

Lejeune, F., and Maquat, L.E. (2005). Mechanistic links between nonsense-mediated mRNA decay and pre-mRNA splicing in mammalian cells. Curr Opin Cell Biol *17*, 309-315.

Levanon, E.Y., Eisenberg, E., Yelin, R., Nemzer, S., Hallegger, M., Shemesh, R., Fligelman, Z.Y., Shoshan, A., Pollock, S.R., Sztybel, D., Olshansky, M., Rechavi, G., and Jantsch, M.F. (2004). Systematic identification of abundant A-to-I editing sites in the human transcriptome. Nat Biotechnol *22*, 1001-1005.

Li, C., and Wong, W.H. (2001). Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. Proc Natl Acad Sci U S A *98*, 31-36.

Li, J.B., Levanon, E.Y., Yoon, J.K., Aach, J., Xie, B., Leproust, E., Zhang, K., Gao, Y., and Church, G.M. (2009). Genome-wide identification of human RNA editing sites by parallel DNA capturing and sequencing. Science *324*, 1210-1213.

Liao, W., Hong, S.H., Chan, B.H., Rudolph, F.B., Clark, S.C., and Chan, L. (1999). APOBEC-2, a cardiac- and skeletal muscle-specific member of the cytidine deaminase supergene family. Biochem Biophys Res Commun *260*, 398-404.

Liebl, F.L., and Featherstone, D.E. (2005). Genes involved in Drosophila glutamate receptor expression and localization. BMC Neurosci *6*, 44.

Liu, S.J., and Zukin, R.S. (2007). Ca2+-permeable AMPA receptors in synaptic plasticity and neuronal death. Trends Neurosci *30*, 126-134.

Liu, Y., Emeson, R.B., and Samuel, C.E. (1999). Serotonin-2C receptor pre-mRNA editing in rat brain and in vitro by splice site variants of the interferon-inducible double-stranded RNA-specific adenosine deaminase ADAR1. J Biol Chem *274*, 18351-18358.

Liu, Y., Lei, M., and Samuel, C.E. (2000). Chimeric double-stranded RNA-specific adenosine deaminase ADAR1 proteins reveal functional selectivity of double-stranded RNA-binding domains from ADAR1 and protein kinase PKR. Proc Natl Acad Sci U S A *97*, 12541-12546.

Lomeli, H., Mosbacher, J., Melcher, T., Hoger, T., Geiger, J.R., Kuner, T., Monyer, H., Higuchi, M., Bach, A., and Seeburg, P.H. (1994). Control of kinetic properties of AMPA receptor channels by nuclear RNA editing. Science *266*, 1709-1713.

Ma, E., Gu, X.Q., Wu, X., Xu, T., and Haddad, G.G. (2001). Mutation in pre-mRNA adenosine deaminase markedly attenuates neuronal tolerance to O2 deprivation in Drosophila melanogaster. J Clin Invest *107*, 685-693.

Maas, S., Melcher, T., Herb, A., Seeburg, P.H., Keller, W., Krause, S., Higuchi, M., and O'Connell, M.A. (1996). Structural requirements for RNA editing in glutamate receptor pre-mRNAs by recombinant double-stranded RNA adenosine deaminase. J Biol Chem *271*, 12221-12226.

Maas, S., Patt, S., Schrey, M., and Rich, A. (2001). Underediting of glutamate receptor GluR-B mRNA in malignant gliomas. Proc Natl Acad Sci U S A *98*, 14687-14692.

Macbeth, M.R., Lingam, A.T., and Bass, B.L. (2004). Evidence for auto-inhibition by the N terminus of hADAR2 and activation by dsRNA binding. Rna *10*, 1563-1571.

Macbeth, M.R., Schubert, H.L., Vandemark, A.P., Lingam, A.T., Hill, C.P., and Bass, B.L. (2005). Inositol hexakisphosphate is bound in the ADAR2 core and required for RNA editing. Science *309*, 1534-1539.

MacDuff, D.A., Demorest, Z.L., and Harris, R.S. (2009). AID can restrict L1 retrotransposition suggesting a dual role in innate and adaptive immunity. Nucleic Acids Res *37*, 1854-1867.

Maizels, N. (2005). Immunoglobulin gene diversification. Annu Rev Genet *39*, 23-46.

Maksakova, I.A., Mager, D.L., and Reiss, D. (2008). Keeping active endogenous retroviral-like elements in check: the epigenetic perspective. Cell Mol Life Sci *65*, 3329-3347.

Mandel, C.R., Bai, Y., and Tong, L. (2008). Protein factors in pre-mRNA 3'-end processing. Cell Mol Life Sci *65*, 1099-1122.

Mangeat, B., Turelli, P., Liao, S., and Trono, D. (2004). A single amino acid determinant governs the species-specific sensitivity of APOBEC3G to Vif action. J Biol Chem *279*, 14481-14483.

Maratou, K., Forster, T., Costa, Y., Taggart, M., Speed, R.M., Ireland, J., Teague, P., Roy, D., and Cooke, H.J. (2004). Expression profiling of the developing testis in wild-type and Dazl knockout mice. Mol Reprod Dev *67*, 26-54.

Marcucci, R., Romano, M., Feiguin, F., O'Connell, M.A., and Baralle, F.E. (2009). Dissecting the splicing mechanism of the Drosophila editing enzyme; dADAR. Nucleic Acids Res *37*, 1663-1671.

Mariani, R., Chen, D., Schrofelbauer, B., Navarro, F., Konig, R., Bollman, B., Munk, C., Nymark-McMahon, H., and Landau, N.R. (2003). Species-specific exclusion of APOBEC3G from HIV-1 virions by Vif. Cell *114*, 21-31.

Marsh, J.L., Walker, H., Theisen, H., Zhu, Y.Z., Fielder, T., Purcell, J., and Thompson, L.M. (2000). Expanded polyglutamine peptides alone are intrinsically cytotoxic and cause neurodegeneration in Drosophila. Hum Mol Genet *9*, 13-25.

Marston, F.A. (1986). The purification of eukaryotic polypeptides synthesized in Escherichia coli. Biochem J *240*, 1-12.

Marzluff, W.F., Wagner, E.J., and Duronio, R.J. (2008). Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. Nat Rev Genet *9*, 843-854.

McCracken, S., Fong, N., Rosonina, E., Yankulov, K., Brothers, G., Siderovski, D., Hessel, A., Foster, S., Shuman, S., and Bentley, D.L. (1997a). 5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. Genes Dev *11*, 3306-3318.

McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S.D., Wickens, M., and Bentley, D.L. (1997b). The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. Nature *385*, 357-361.

McGurk. L. (2008). *Drosophila* lacking adenosine to inosine RNA editing. (The University of Edinburgh).

Mehta, A., Kinter, M.T., Sherman, N.E., and Driscoll, D.M. (2000). Molecular cloning of apobec-1 complementation factor, a novel RNA-binding protein involved in the editing of apolipoprotein B mRNA. Mol Cell Biol *20*, 1846-1854.

Melcher, T., Maas, S., Herb, A., Sprengel, R., Higuchi, M., and Seeburg, P.H. (1996a). RED2, a brain-specific member of the RNA-specific adenosine deaminase family. J Biol Chem *271*, 31795-31798.

Melcher, T., Maas, S., Herb, A., Sprengel, R., Seeburg, P.H., and Higuchi, M. (1996b). A mammalian RNA editing enzyme. Nature *379*, 460-464.

Mikl, M.C., Watt, I.N., Lu, M., Reik, W., Davies, S.L., Neuberger, M.S., and Rada, C. (2005). Mice deficient in APOBEC2 and APOBEC3. Mol Cell Biol *25*, 7270-7277.

Miller, D., Ostermeier, G.C., and Krawetz, S.A. (2005). The controversy, potential and roles of spermatozoal RNA. Trends Mol Med *11*, 156-163.

Min, K.T., and Benzer, S. (1999). Preventing neurodegeneration in the Drosophila mutant bubblegum. Science *284*, 1985-1988.

Miyamura, Y., Suzuki, T., Kono, M., Inagaki, K., Ito, S., Suzuki, N., and Tomita, Y. (2003). Mutations of the RNA-specific adenosine deaminase gene (DSRAD) are involved in dyschromatosis symmetrica hereditaria. Am J Hum Genet *73*, 693-699.

Moore, M.J., and Sharp, P.A. (1993). Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. Nature *365*, 364-368.

Moore, M.J., and Silver, P.A. (2008). Global analysis of mRNA splicing. Rna *14*, 197-203.

Morgan, H.D., Dean, W., Coker, H.A., Reik, W., and Petersen-Mahrt, S.K. (2004). Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. J Biol Chem *279*, 52353-52360.

Morgan, H.D., Santos, F., Green, K., Dean, W., and Reik, W. (2005). Epigenetic reprogramming in mammals. Hum Mol Genet *14 Spec No 1*, R47-58.

Morrison, J.R., Paszty, C., Stevens, M.E., Hughes, S.D., Forte, T., Scott, J., and Rubin, E.M. (1996). Apolipoprotein B RNA editing enzyme-deficient mice are viable despite alterations in lipoprotein metabolism. Proc Natl Acad Sci U S A *93*, 7154-7159.

Morse, D.P., Aruscavage, P.J., and Bass, B.L. (2002). RNA hairpins in noncoding regions of human brain and Caenorhabditis elegans mRNA are edited by adenosine deaminases that act on RNA. Proc Natl Acad Sci U S A *99*, 7906-7911.

Morse, D.P., and Bass, B.L. (1999). Long RNA hairpins that contain inosine are present in Caenorhabditis elegans poly(A)+ RNA. Proc Natl Acad Sci U S A *96*, 6048-6053.

Mount, S.M., and Salz, H.K. (2000). Pre-messenger RNA processing factors in the Drosophila genome. J Cell Biol *150*, F37-44.

Muhlig-Versen, M., da Cruz, A.B., Tschape, J.A., Moser, M., Buttner, R., Athenstaedt, K., Glynn, P., and Kretzschmar, D. (2005). Loss of Swiss cheese/neuropathy target esterase activity causes disruption of phosphatidylcholine homeostasis and neuronal and glial death in adult Drosophila. J Neurosci *25*, 2865-2873.

Muqit, M.M., and Feany, M.B. (2002). Modelling neurodegenerative diseases in Drosophila: a fruitful approach? Nat Rev Neurosci *3*, 237-243.

Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., and Honjo, T. (2000). Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. Cell *102*, 553-563.

Muto, T., Okazaki, I.M., Yamada, S., Tanaka, Y., Kinoshita, K., Muramatsu, M., Nagaoka, H., and Honjo, T. (2006). Negative regulation of activation-induced cytidine deaminase in B cells. Proc Natl Acad Sci U S A *103*, 2752-2757.

Nakamuta, M., Chang, B.H., Zsigmond, E., Kobayashi, K., Lei, H., Ishida, B.Y., Oka, K., Li, E., and Chan, L. (1996). Complete phenotypic characterization of apobec-1 knockout mice with a wild-type genetic background and a human apolipoprotein B transgenic background, and restoration of apolipoprotein B mRNA editing by somatic gene transfer of Apobec-1. J Biol Chem *271*, 25981-25988.

Nambu, Y., Sugai, M., Gonda, H., Lee, C.G., Katakai, T., Agata, Y., Yokota, Y., and Shimizu, A. (2003). Transcription-coupled events associating with immunoglobulin switch region chromatin. Science *302*, 2137-2140.

Navaratnam, N., Morrison, J.R., Bhattacharya, S., Patel, D., Funahashi, T., Giannoni, F., Teng, B.B., Davidson, N.O., and Scott, J. (1993). The p27 catalytic subunit of the apolipoprotein B mRNA editing enzyme is a cytidine deaminase. J Biol Chem *268*, 20709-20712.

Niswender, C.M., Herrick-Davis, K., Dilley, G.E., Meltzer, H.Y., Overholser, J.C., Stockmeier, C.A., Emeson, R.B., and Sanders-Bush, E. (2001). RNA editing of the human serotonin 5-HT2C receptor. alterations in suicide and implications for serotonergic pharmacotherapy. Neuropsychopharmacology *24*, 478-491.

Nonaka, T., Doi, T., Toyoshima, T., Muramatsu, M., Honjo, T., and Kinoshita, K. (2009). Carboxy-terminal domain of AID required for its mRNA complex formation in vivo. Proc Natl Acad Sci U S A *106*, 2747-2751.

O'Connell, M.A., and Keller, W. (1994). Purification and properties of double-stranded RNA-specific adenosine deaminase from calf thymus. Proc Natl Acad Sci U S A *91*, 10596-10600.

O'Connell, M.A., Krause, S., Higuchi, M., Hsuan, J.J., Totty, N.F., Jenny, A., and Keller, W. (1995). Cloning of cDNAs encoding mammalian double-stranded RNA-specific adenosine deaminase. Mol Cell Biol *15*, 1389-1397.

Ohlson, J., Enstero, M., Sjoberg, B.M., and Ohman, M. (2005). A method to find tissue-specific novel sites of selective adenosine deamination. Nucleic Acids Res *33*, e167.

Ohman, M., Kallman, A.M., and Bass, B.L. (2000). In vitro analysis of the binding of ADAR2 to the pre-mRNA encoding the GluR-B R/G site. Rna *6*, 687-697.

Okada, Y., Yamagata, K., Hong, K., Wakayama, T., and Zhang, Y. (2010). A role for the elongator complex in zygotic paternal genome demethylation. Nature.

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247-257.

Okeoma, C.M., Lovsin, N., Peterlin, B.M., and Ross, S.R. (2007). APOBEC3 inhibits mouse mammary tumour virus replication in vivo. Nature *445*, 927-930.

Olson, S., Blanchette, M., Park, J., Savva, Y., Yeo, G.W., Yeakley, J.M., Rio, D.C., and Graveley, B.R. (2007). A regulator of *Dscam* mutually exclusive splicing fidelity. Nature Structural & Molecular Biology *14*, 1134-1140.

Opi, S., Takeuchi, H., Kao, S., Khan, M.A., Miyagi, E., Goila-Gaur, R., Iwatani, Y., Levin, J.G., and Strebel, K. (2006). Monomeric APOBEC3G is catalytically active and has antiviral activity. J Virol *80*, 4673-4682.

Ostermeier, G.C., Miller, D., Huntriss, J.D., Diamond, M.P., and Krawetz, S.A. (2004). Reproductive biology: delivering spermatozoan RNA to the oocyte. Nature *429*, 154.

Palladino, M.J., Keegan, L.P., O'Connell, M.A., and Reenan, R.A. (2000a). dADAR, a Drosophila double-stranded RNA-specific adenosine deaminase is highly developmentally regulated and is itself a target for RNA editing. Rna *6*, 1004-1018.

Palladino, M.J., Keegan, L.P., O'Connell, M.A., and Reenan, R.A. (2000b). A-to-I pre-mRNA editing in Drosophila is primarily involved in adult nervous system function and integrity. Cell *102*, 437-449.

Paraskevopoulou, C., Fairhurst, S.A., Lowe, D.J., Brick, P., and Onesti, S. (2006). The Elongator subunit Elp3 contains a Fe4S4 cluster and binds S-adenosylmethionine. Mol Microbiol *59*, 795-806.

Parkes, T.L., Elia, A.J., Dickinson, D., Hilliker, A.J., Phillips, J.P., and Boulianne, G.L. (1998). Extension of Drosophila lifespan by overexpression of human SOD1 in motorneurons. Nat Genet *19*, 171-174.

Patel, S.R., and Saide, J.D. (2005). Stretchin-klp, a novel Drosophila indirect flight muscle protein, has both myosin dependent and independent isoforms. J Muscle Res Cell Motil *26*, 213-224.

Patterson, J.B., and Samuel, C.E. (1995). Expression and regulation by interferon of a double-stranded-RNA-specific adenosine deaminase from human cells: evidence for two forms of the deaminase. Mol Cell Biol *15*, 5376-5388.

Patton, D.E., Silva, T., and Bezanilla, F. (1997). RNA editing generates a diverse array of transcripts encoding squid Kv2 K+ channels with altered functional properties. Neuron *19*, 711-722.

Paul, M.S., and Bass, B.L. (1998). Inosine exists in mRNA at tissue-specific levels and is most abundant in brain mRNA. Embo J *17*, 1120-1127.

Paz, N., Levanon, E.Y., Amariglio, N., Heimberger, A.B., Ram, Z., Constantini, S., Barbash, Z.S., Adamsky, K., Safran, M., Hirschberg, A., Krupsky, M., Ben-Dov, I., Cazacu, S., Mikkelsen, T., Brodie, C., Eisenberg, E., and Rechavi, G. (2007). Altered adenosine-to-inosine RNA editing in human cancer. Genome Res *17*, 1586-1595.

Peixoto, A.A., and Hall, J.C. (1998). Analysis of temperature-sensitive mutants reveals new genes involved in the courtship song of Drosophila. Genetics *148*, 827-838.

Peng, P.L., Zhong, X., Tu, W., Soundarapandian, M.M., Molner, P., Zhu, D., Lau, L., Liu, S., Liu, F., and Lu, Y. (2006). ADAR2-dependent RNA editing of AMPA receptor subunit GluR2 determines vulnerability of neurons in forebrain ischemia. Neuron *49*, 719-733.

Peters, N.T., Rohrbach, J.A., Zalewski, B.A., Byrkett, C.M., and Vaughn, J.C. (2003). RNA editing and regulation of Drosophila 4f-rnp expression by sas-10 antisense readthrough mRNA transcripts. Rna *9*, 698-710.

Petersen-Mahrt, S.K., Harris, R.S., and Neuberger, M.S. (2002). AID mutates E. coli suggesting a DNA deamination mechanism for antibody diversification. Nature *418*, 99-103.

Petersen-Mahrt, S.K., and Neuberger, M.S. (2003). In vitro deamination of cytosine to uracil in single-stranded DNA by apolipoprotein B editing complex catalytic subunit 1 (APOBEC1). J Biol Chem *278*, 19583-19586.

Petschek, J.P., Mermer, M.J., Scheckelhoff, M.R., Simone, A.A., and Vaughn, J.C. (1996). RNA editing in Drosophila 4f-rnp gene nuclear transcripts by multiple A-to-G conversions. J Mol Biol *259*, 885-890.

Pham, P., Bransteitter, R., Petruska, J., and Goodman, M.F. (2003). Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. Nature *424*, 103-107.

Polson, A.G., Bass, B.L., and Casey, J.L. (1996). RNA editing of hepatitis delta virus antigenome by dsRNA-adenosine deaminase. Nature *380*, 454-456.

Polson, A.G., Ley, H.L., 3rd, Bass, B.L., and Casey, J.L. (1998). Hepatitis delta virus RNA editing is highly specific for the amber/W site and is suppressed by hepatitis delta antigen. Mol Cell Biol *18*, 1919-1926.

Poltoratsky, V.P., Wilson, S.H., Kunkel, T.A., and Pavlov, Y.I. (2004). Recombinogenic phenotype of human activation-induced cytosine deaminase. J Immunol *172*, 4308-4313.

Popp, C., Dean, W., Feng, S., Cokus, S.J., Andrews, S., Pellegrini, M., Jacobsen, S.E., and Reik, W. (2010). Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. Nature *463*, 1101-1105.

Poulsen, H., Jorgensen, R., Heding, A., Nielsen, F.C., Bonven, B., and Egebjerg, J. (2006). Dimerization of ADAR2 is mediated by the double-stranded RNA binding domain. Rna *12*, 1350-1360.

Poulsen, H., Nilsson, J., Damgaard, C.K., Egebjerg, J., and Kjems, J. (2001). CRM1 mediates the export of ADAR1 through a nuclear export signal within the Z-DNA binding domain. Mol Cell Biol *21*, 7862-7871.

Powell, L.M., Wallis, S.C., Pease, R.J., Edwards, Y.H., Knott, T.J., and Scott, J. (1987). A novel form of tissue-specific RNA processing produces apolipoprotein-48 in intestine. Cell *50*, 831-840.

Prasanth, K.V., Prasanth, S.G., Xuan, Z., Hearn, S., Freier, S.M., Bennett, C.F., Zhang, M.Q., and Spector, D.L. (2005). Regulating gene expression through RNA nuclear retention. Cell *123*, 249-263.

Prochnow, C., Bransteitter, R., Klein, M.G., Goodman, M.F., and Chen, X.S. (2007). The APOBEC-2 crystal structure and functional implications for the deaminase AID. Nature *445*, 447-451.

Proudfoot, N. (2004). New perspectives on connecting messenger RNA 3' end formation to transcription. Curr Opin Cell Biol *16*, 272-278.

Qin, G., Schwarz, T., Kittel, R.J., Schmid, A., Rasse, T.M., Kappei, D., Ponimaskin, E., Heckmann, M., and Sigrist, S.J. (2005). Four different subunits are essential for expressing the synaptic glutamate receptor at neuromuscular junctions of Drosophila. J Neurosci *25*, 3209-3218.

Rada, C., Di Noia, J.M., and Neuberger, M.S. (2004). Mismatch recognition and uracil excision provide complementary paths to both Ig switching and the A/T-focused phase of somatic mutation. Mol Cell *16*, 163-171.

Rai, K., Huggins, I.J., James, S.R., Karpf, A.R., Jones, D.A., and Cairns, B.R. (2008). DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. Cell *135*, 1201-1212.

Raitskin, O., Cho, D.S., Sperling, J., Nishikura, K., and Sperling, R. (2001). RNA editing activity is associated with splicing factors in lnRNP particles: The nuclear pre-mRNA processing machinery. Proc Natl Acad Sci U S A *98*, 6571-6576.

Rajeevan, M.S., Vernon, S.D., Taysavang, N., and Unger, E.R. (2001). Validation of array-based gene expression profiles by real-time (kinetic) RT-PCR. J Mol Diagn *3*, 26-31.

Ramiro, A.R., Nussenzweig, M.C., and Nussenzweig, A. (2006). Switching on chromosomal translocations. Cancer Res *66*, 7837-7839.

Ramos, A., Grunert, S., Adams, J., Micklem, D.R., Proctor, M.R., Freund, S., Bycroft, M., St Johnston, D., and Varani, G. (2000). RNA recognition by a Staufen double-stranded RNA-binding domain. Embo J *19*, 997-1009.

Ranganayakulu, G., Schulz, R.A., and Olson, E.N. (1996). Wingless signaling induces nautilus expression in the ventral mesoderm of the Drosophila embryo. Dev Biol *176*, 143-148.

Reenan, R.A. (2005). Molecular determinants and guided evolution of species-specific RNA editing. Nature *434*, 409-413.

Reenan, R.A., Hanrahan, C.J., and Ganetzky, B. (2000). The mle(napts) RNA helicase mutation in drosophila results in a splicing catastrophe of the para Na+ channel transcript in a region of RNA editing. Neuron *25*, 139-149.

Retelska, D., Iseli, C., Bucher, P., Jongeneel, C.V., and Naef, F. (2006). Similarities and differences of polyadenylation signals in human and fly. BMC Genomics *7*, 176.

Revy, P., Muto, T., Levy, Y., Geissmann, F., Plebani, A., Sanal, O., Catalan, N., Forveille, M., Dufourcq-Labelouse, R., Gennery, A., Tezcan, I., Ersoy, F., Kayserili, H., Ugazio, A.G., Brousse, N., Muramatsu, M., Notarangelo, L.D., Kinoshita, K., Honjo, T., Fischer, A., and Durandy, A. (2000). Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). Cell *102*, 565-575.

Ring, G.M., O'Connell, M.A., and Keegan, L.P. (2004). Purification and assay of recombinant ADAR proteins expressed in the yeast Pichia pastoris or in Escherichia coli. Methods Mol Biol *265*, 219-238.

Robbiani, D.F., Bunting, S., Feldhahn, N., Bothmer, A., Camps, J., Deroubaix, S., McBride, K.M., Klein, I.A., Stone, G., Eisenreich, T.R., Ried, T., Nussenzweig, A., and Nussenzweig, M.C. (2009). AID produces DNA double-strand breaks in non-Ig genes and mature B cell lymphomas with reciprocal chromosome translocations. Mol Cell *36*, 631-641.

Roberts, R.J., and Cheng, X. (1998). Base flipping. Annu Rev Biochem *67*, 181-198.

Rogers, R.L., Bedford, T., and Hartl, D.L. (2009). Formation and longevity of chimeric and duplicate genes in Drosophila melanogaster. Genetics *181*, 313-322.

Rogozin, I.B., Basu, M.K., Jordan, I.K., Pavlov, Y.I., and Koonin, E.V. (2005). APOBEC4, a new member of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases predicted by computational analysis. Cell Cycle *4*, 1281-1285.

Rousseaux, S., Caron, C., Govin, J., Lestrat, C., Faure, A.K., and Khochbin, S. (2005). Establishment of male-specific epigenetic information. Gene *345*, 139-153.

Rubio, M.A., Pastar, I., Gaston, K.W., Ragone, F.L., Janzen, C.J., Cross, G.A., Papavasiliou, F.N., and Alfonzo, J.D. (2007). An adenosine-to-inosine tRNA-editing enzyme that can perform C-to-U deamination of DNA. Proc Natl Acad Sci U S A *104*, 7821-7826.

Rueter, S.M., Dawson, T.R., and Emeson, R.B. (1999). Regulation of alternative splicing by RNA editing. Nature *399*, 75-80.

Ryman, K., Fong, N., Bratt, E., Bentley, D.L., and Ohman, M. (2007). The C-terminal domain of RNA Pol II helps ensure that editing precedes splicing of the GluR-B transcript. Rna *13*, 1071-1078.

Sansam, C.L., Wells, K.S., and Emeson, R.B. (2003). Modulation of RNA editing by functional nucleolar sequestration of ADAR2. Proc Natl Acad Sci U S A *100*, 14018-14023.

Santiago, M.L., Montano, M., Benitez, R., Messer, R.J., Yonemoto, W., Chesebro, B., Hasenkrug, K.J., and Greene, W.C. (2008). Apobec3 encodes Rfv3, a gene influencing neutralizing antibody control of retrovirus infection. Science *321*, 1343-1346.

Santos, F., Hendrich, B., Reik, W., and Dean, W. (2002). Dynamic reprogramming of DNA methylation in the early mouse embryo. Dev Biol *241*, 172-182.

Sato, Y., Probst, H.C., Tatsumi, R., Ikeuchi, Y., Neuberger, M.S., and Rada, C. (2009). Deficiency in APOBEC2 leads to a shift in muscle fiber-type, diminished body mass and myopathy. J Biol Chem.

Sattelle, D.B., Jones, A.K., Sattelle, B.M., Matsuda, K., Reenan, R., and Biggin, P.C. (2005). Edit, cut and paste in the nicotinic acetylcholine receptor gene family of Drosophila melanogaster. Bioessays *27*, 366-376.

Sawyer, S.L., Emerman, M., and Malik, H.S. (2004). Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. PLoS Biol *2*, E275.

Scadden, A.D. (2005). The RISC subunit Tudor-SN binds to hyper-edited double-stranded RNA and promotes its cleavage. Nat Struct Mol Biol *12*, 489-496.

Scadden, A.D. (2007). Inosine-containing dsRNA binds a stress-granule-like complex and downregulates gene expression in trans. Mol Cell *28*, 491-500.

Scadden, A.D., and O'Connell, M.A. (2005). Cleavage of dsRNAs hyper-edited by ADARs occurs at preferred editing sites. Nucleic Acids Res *33*, 5954-5964.

Schmucker, D. (2007). Molecular diversity of Dscam: recognition of molecular identity in neuronal wiring. Nat Rev Neurosci *8*, 915-920.

Schmucker, D., Clemens, J.C., Shu, H., Worby, C.A., Xiao, J., Muda, M., Dixon, J.E., and Zipursky, S.L. (2000). Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. Cell *101*, 671-684.

Schoft, V.K., Schopoff, S., and Jantsch, M.F. (2007). Regulation of glutamate receptor B pre-mRNA splicing by RNA editing. Nucleic Acids Res *35*, 3723-3732.

Schreck, S., Buettner, M., Kremmer, E., Bogdan, M., Herbst, H., and Niedobitek, G. (2006). Activation-induced cytidine deaminase (AID) is expressed in normal spermatogenesis but only infrequently in testicular germ cell tumours. J Pathol *210*, 26-31.

Schrofelbauer, B., Chen, D., and Landau, N.R. (2004). A single amino acid of APOBEC3G controls its species-specific interaction with virion infectivity factor (Vif). Proc Natl Acad Sci U S A *101*, 3927-3932.

Schumacher, A.J., Nissley, D.V., and Harris, R.S. (2005). APOBEC3G hypermutates genomic DNA and inhibits Ty1 retrotransposition in yeast. Proc Natl Acad Sci U S A *102*, 9854-9859.

Schumacher, J.M., Lee, K., Edelhoff, S., and Braun, R.E. (1995). Distribution of Tenr, an RNA-binding protein, in a lattice-like network within the spermatid nucleus in the mouse. Biol Reprod *52*, 1274-1283.

Semenov, E.P., and Pak, W.L. (1999). Diversification of Drosophila chloride channel gene by multiple posttranscriptional mRNA modifications. J Neurochem *72*, 66-72.

Serra, M.J., Smolter, P.E., and Westhof, E. (2004). Pronounced instability of tandem IU base pairs in RNA. Nucleic Acids Res *32*, 1824-1828.

Shan, X., Chang, Y., and Lin, C.L. (2007). Messenger RNA oxidation is an early event preceding cell death and causes reduced protein expression. Faseb J *21*, 2753-2764.

Sharma, P.M., Bowman, M., Madden, S.L., Rauscher, F.J., 3rd, and Sukumar, S. (1994). RNA editing in the Wilms' tumor susceptibility gene, WT1. Genes Dev *8*, 720-731.

Sheehy, A.M., Gaddis, N.C., Choi, J.D., and Malim, M.H. (2002). Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. Nature *418*, 646-650.

Simonsen, A., Cumming, R.C., Lindmo, K., Galaviz, V., Cheng, S., Rusten, T.E., and Finley, K.D. (2007). Genetic modifiers of the Drosophila blue cheese gene link defects in lysosomal transport with decreased life span and altered ubiquitinated-protein profiles. Genetics *176*, 1283-1297.

Singh, M., Kesterson, R.A., Jacobs, M.M., Joers, J.M., Gore, J.C., and Emeson, R.B. (2007). Hyperphagia-mediated obesity in transgenic mice misexpressing the RNA-editing enzyme ADAR2. J Biol Chem *282*, 22448-22459.

Smith, H.C. (2007). Measuring editing activity and identifying cytidine-to-uridine mRNA editing factors in cells and biochemical isolates. Methods Enzymol *424*, 389-416.

Smith, L.A., Wang, X., Peixoto, A.A., Neumann, E.K., Hall, L.M., and Hall, J.C. (1996). A Drosophila calcium channel alpha1 subunit gene maps to a genetic locus associated with behavioral and visual defects. J Neurosci *16*, 7868-7879.

Sodhi, M.S., Burnet, P.W., Makoff, A.J., Kerwin, R.W., and Harrison, P.J. (2001). RNA editing of the 5-HT(2C) receptor is reduced in schizophrenia. Mol Psychiatry *6*, 373-379.

Sommer, B., Keinanen, K., Verdoorn, T.A., Wisden, W., Burnashev, N., Herb, A., Kohler, M., Takagi, T., Sakmann, B., and Seeburg, P.H. (1990). Flip and flop: a cell-specific functional switch in glutamate-operated channels of the CNS. Science *249*, 1580-1585.

Sowden, M.P., Ballatori, N., Jensen, K.L., Reed, L.H., and Smith, H.C. (2002). The editosome for cytidine to uridine mRNA editing has a native complexity of 27S: identification of intracellular domains containing active and inactive editing factors. J Cell Sci *115*, 1027-1039.

Stapleton, M., Carlson, J.W., and Celniker, S.E. (2006). RNA editing in Drosophila melanogaster: New targets and functional consequences. Rna *12*, 1922-1932.

Stefl, R., Xu, M., Skrisovska, L., Emeson, R.B., and Allain, F.H. (2006). Structure and specific RNA binding of ADAR2 double-stranded RNA binding motifs. Structure *14*, 345-355.

Stenglein, M.D., Burns, M.B., Li, M., Lengyel, J., and Harris, R.S. (2010). APOBEC3 proteins mediate the clearance of foreign DNA from human cells. Nat Struct Mol Biol.

Stenglein, M.D., and Harris, R.S. (2006). APOBEC3B and APOBEC3F inhibit L1 retrotransposition by a DNA deamination-independent mechanism. J Biol Chem *281*, 16837-16841.

Stephens, O.M., Haudenschild, B.L., and Beal, P.A. (2004). The binding selectivity of ADAR2's dsRBMs contributes to RNA-editing selectivity. Chem Biol *11*, 1239-1250.

Suspene, R., Guetard, D., Henry, M., Sommer, P., Wain-Hobson, S., and Vartanian, J.P. (2005). Extensive editing of both hepatitis B virus DNA strands by APOBEC3 cytidine deaminases in vitro and in vivo. Proc Natl Acad Sci U S A *102*, 8321-8326.

Swank, D.M., Wells, L., Kronert, W.A., Morrill, G.E., and Bernstein, S.I. (2000). Determining structure/function relationships for sarcomeric myosin heavy chain by genetic and transgenic manipulation of Drosophila. Microsc Res Tech *50*, 430-442.

Tam, W., Ben-Yehuda, D., and Hayward, W.S. (1997). bic, a novel gene activated by proviral insertions in avian leukosis virus-induced lymphomas, is likely to function through its noncoding RNA. Mol Cell Biol *17*, 1490-1502.

Tanaka, M., Chock, P.B., and Stadtman, E.R. (2007). Oxidized messenger RNA induces translation errors. Proc Natl Acad Sci U S A *104*, 66-71.

Teng, B., Burant, C.F., and Davidson, N.O. (1993). Molecular cloning of an apolipoprotein B messenger RNA editing protein. Science *260*, 1816-1819.

Teng, G., Hakimpour, P., Landgraf, P., Rice, A., Tuschl, T., Casellas, R., and Papavasiliou, F.N. (2008). MicroRNA-155 is a negative regulator of activation-induced cytidine deaminase. Immunity *28*, 621-629.

Thai, T.H., Calado, D.P., Casola, S., Ansel, K.M., Xiao, C., Xue, Y., Murphy, A., Frendewey, D., Valenzuela, D., Kutok, J.L., Schmidt-Supprian, M., Rajewsky, N., Yancopoulos, G., Rao, A., and Rajewsky, K. (2007). Regulation of the germinal center response by microRNA-155. Science *316*, 604-608.

Tonkin, L.A., and Bass, B.L. (2003). Mutations in RNAi rescue aberrant chemotaxis of ADAR mutants. Science *302*, 1725.

Towers, P.R., and Sattelle, D.B. (2002). A Drosophila melanogaster cell line (S2) facilitates post-genome functional analysis of receptors and ion channels. Bioessays *24*, 1066-1073.

Turelli, P., Mangeat, B., Jost, S., Vianin, S., and Trono, D. (2004). Inhibition of hepatitis B virus replication by APOBEC3G. Science *303*, 1829.

Ule, J., Jensen, K.B., Ruggiu, M., Mele, A., Ule, A., and Darnell, R.B. (2003). CLIP identifies Nova-regulated RNA networks in the brain. Science *302*, 1212-1215.

Ule, J., Ule, A., Spencer, J., Williams, A., Hu, J.S., Cline, M., Wang, H., Clark, T., Fraser, C., Ruggiu, M., Zeeberg, B.R., Kane, D., Weinstein, J.N., Blume, J., and Darnell, R.B. (2005). Nova regulates brain-specific splicing to shape the synapse. Nat Genet *37*, 844-852.

Valente, L., and Nishikura, K. (2007). RNA binding-independent dimerization of adenosine deaminases acting on RNA and dominant negative effects of nonfunctional subunits on dimer functions. J Biol Chem *282*, 16054-16061.

Wagner, E.J., Burch, B.D., Godfrey, A.C., Salzler, H.R., Duronio, R.J., and Marzluff, W.F. (2007). A genome-wide RNA interference screen reveals that variant histones are necessary for replication-dependent histone pre-mRNA processing. Mol Cell *28*, 692-699.

Walton, M., Sirimanne, E., Williams, C., Gluckman, P., and Dragunow, M. (1996). The role of the cyclic AMP-responsive element binding protein (CREB) in hypoxic-ischemic brain damage and repair. Brain Res Mol Brain Res *43*, 21-29.

Wang, Q., Miyakoda, M., Yang, W., Khillan, J., Stachura, D.L., Weiss, M.J., and Nishikura, K. (2004). Stress-induced apoptosis associated with null mutation of ADAR1 RNA editing deaminase gene. J Biol Chem *279*, 4952-4961.

Wang, X. (2008). miRDB: a microRNA target prediction and functional annotation database with a wiki interface. Rna *14*, 1012-1017.

Wang, X., and El Naqa, I.M. (2008). Prediction of both conserved and nonconserved microRNA targets in animals. Bioinformatics *24*, 325-332.

Wang, Z., Tollervey, J., Briese, M., Turner, D., and Ule, J. (2009). CLIP: construction of cDNA libraries for high-throughput sequencing from RNAs cross-linked to proteins in vivo. Methods *48*, 287-293.

Watanabe, T., Takeda, A., Tsukiyama, T., Mise, K., Okuno, T., Sasaki, H., Minami, N., and Imai, H. (2006). Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes. Genes Dev *20*, 1732-1743.

Wedekind, J.E., Gillilan, R., Janda, A., Krucinska, J., Salter, J.D., Bennett, R.P., Raina, J., and Smith, H.C. (2006). Nanostructures of APOBEC3G support a hierarchical assembly model of high molecular mass ribonucleoprotein particles from dimeric subunits. J Biol Chem *281*, 38122-38126.

Wells, L., Edwards, K.A., and Bernstein, S.I. (1996). Myosin heavy chain isoforms regulate muscle function but not myofibril assembly. Embo J *15*, 4454-4459.

Wichroski, M.J., Robb, G.B., and Rana, T.M. (2006). Human retroviral host restriction factors APOBEC3G and APOBEC3F localize to mRNA processing bodies. PLoS Pathog *2*, e41.

Wilson, D.K., Rudolph, F.B., and Quiocho, F.A. (1991). Atomic structure of adenosine deaminase complexed with a transition-state analog: understanding catalysis and immunodeficiency mutations. Science *252*, 1278-1284.

Wong, S.K., Sato, S., and Lazinski, D.W. (2001). Substrate recognition by ADAR1 and ADAR2. Rna *7*, 846-858.

Xia, S., Miyashita, T., Fu, T.F., Lin, W.Y., Wu, C.L., Pyzocha, L., Lin, I.R., Saitoe, M., Tully, T., and Chiang, A.S. (2005a). NMDA receptors mediate olfactory learning and memory in Drosophila. Curr Biol *15*, 603-615.

Xia, S., Yang, J., Su, Y., Qian, J., Ma, E., and Haddad, G.G. (2005b). Identification of new targets of Drosophila pre-mRNA adenosine deaminase. Physiol Genomics *20*, 195-202.

Xu, H., Svarovskaia, E.S., Barr, R., Zhang, Y., Khan, M.A., Strebel, K., and Pathak, V.K. (2004). A single amino acid substitution in human APOBEC3G antiretroviral enzyme confers resistance to HIV-1 virion infectivity factor-induced depletion. Proc Natl Acad Sci U S A *101*, 5652-5657.

XuFeng, R., Boyer, M.J., Shen, H., Li, Y., Yu, H., Gao, Y., Yang, Q., Wang, Q., and Cheng, T. (2009). ADAR1 is required for hematopoietic progenitor cell survival via RNA editing. Proc Natl Acad Sci U S A *106*, 17763-17768.

Yamanaka, S., Balestra, M.E., Ferrell, L.D., Fan, J., Arnold, K.S., Taylor, S., Taylor, J.M., and Innerarity, T.L. (1995). Apolipoprotein B mRNA-editing protein induces hepatocellular carcinoma and dysplasia in transgenic animals. Proc Natl Acad Sci U S A *92*, 8483-8487.

Yang, B., Chen, K., Zhang, C., Huang, S., and Zhang, H. (2007). Virion-associated uracil DNA glycosylase-2 and apurinic/apyrimidinic endonuclease are involved in the degradation of APOBEC3G-edited nascent HIV-1 DNA. J Biol Chem *282*, 11667-11675.

Yang, L., Zhao, L., Gan, Z., He, Z., Xu, J., Gao, X., Wang, X., Han, W., Chen, L., Xu, T., Li, W., and Liu, Y. (2010). Deficiency in RNA editing enzyme ADAR2 impairs regulated exocytosis. Faseb J.

Yang, W., Chendrimada, T.P., Wang, Q., Higuchi, M., Seeburg, P.H., Shiekhattar, R., and Nishikura, K. (2006). Modulation of microRNA processing and expression through RNA editing by ADAR deaminases. Nat Struct Mol Biol *13*, 13-21.

Yasuyama, K., Kitamoto, T., and Salvaterra, P.M. (1995). Immunocytochemical study of choline acetyltransferase in Drosophila melanogaster: an analysis of cis-regulatory regions controlling expression in the brain of cDNA-transformed flies. J Comp Neurol *361*, 25-37.

Yeo, G., Holste, D., Kreiman, G., and Burge, C.B. (2004). Variation in alternative splicing across human tissues. Genome Biol *5*, R74.

Yu, F., Zingler, N., Schumann, G., and Stratling, W.H. (2001). Methyl-CpG-binding protein 2 represses LINE-1 expression and retrotransposition but not Alu transcription. Nucleic Acids Res *29*, 4493-4501.

Yu, Q., Konig, R., Pillai, S., Chiles, K., Kearney, M., Palmer, S., Richman, D., Coffin, J.M., and Landau, N.R. (2004). Single-strand specificity of APOBEC3G accounts for minus-strand deamination of the HIV genome. Nat Struct Mol Biol *11*, 435-442.

Zarrin, A.A., Tian, M., Wang, J., Borjeson, T., and Alt, F.W. (2005). Influence of switch region length on immunoglobulin class switch recombination. Proc Natl Acad Sci U S A *102*, 2466-2470.

Zennou, V., Perez-Caballero, D., Gottlinger, H., and Bieniasz, P.D. (2004). APOBEC3G incorporation into human immunodeficiency virus type 1 particles. J Virol *78*, 12058-12061.

Zhang, B., Egli, D., Georgiev, O., and Schaffner, W. (2001). The Drosophila homolog of mammalian zinc finger factor MTF-1 activates transcription in response to heavy metals. Mol Cell Biol *21*, 4505-4514.

Zhang, H., Liu, J., Li, C.R., Momen, B., Kohanski, R.A., and Pick, L. (2009). Deletion of Drosophila insulin-like peptides causes growth defects and metabolic abnormalities. Proc Natl Acad Sci U S A *106*, 19617-19622.

Zou, S., Meadows, S., Sharp, L., Jan, L.Y., and Jan, Y.N. (2000). Genome-wide study of aging and oxidative stress response in Drosophila melanogaster. Proc Natl Acad Sci U S A *97*, 13726-13731.

Chapter 8: Appendix I

ADAR RNA-binding mutant (ADAR-RRM)

An ADAR RNA-binding mutant termed ADAR-RRM was generated as a negative control for analysis of ADAR RNA-binding. Several mutations within the dsRBDs were introduced by site-directed mutagenesis, followed by sequence analysis to ensure only the specified changes were introduced. These mutations changed basic lysine residues within a conserved motif (KKxxK, where x indicates any amino acid) involved in RNA interactions for either alanine or glutamic acid residues such that the motif is changed to EAxxA. The mutations are described in Valente & Nishikura (2007), where they were used to demonstrate the ADAR proteins which are incapable of binding RNA are capable of forming homodimers (Valente and Nishikura, 2007). The amount of protein produced is not described however an insect cell expression system is used and equivalent amounts of protein were loaded on the immunoblots. These mutations were made in both dsRBD of *Drosophila* ADAR and cloned into different expression vectors, and human ADAR2 containing this mutation was generated for expression in flies. The construct for expression in *Drosophila* had to be injected, balanced and crossed with a GAL4 driver line prior to analysis (Jim Brindle and Liam Keegan) and was therefore only available for use at the end of my PhD.

However during various experiments I observed that the dADAR RRM mutant protein was expressed at a consistently lower level than wild type ADAR protein. The reason for the difference in expression level is unclear as inactive ADAR mutants are usually expressed at a higher level than wild type ADAR (unpublished observation). Therefore the ADAR RRM protein may be unstable.

I observed that during expression and purification of ADAR RRM mutant protein from *Pichia pastoris* that the protein was expressed at a lower level than the wild type ADAR protein, which was expressed and purified alongside the mutant as a control (Figure AI.1). Bradford assay confirmed the difference in protein content was approximately two-fold for the second elution fraction E2 (WT = 186ng/ul, RRM = 100ng/ul). This is unusual as the catalytic site mutant ADAR EA protein is consistently expressed at a higher level than wild type protein (personal observation), leading to the hypothesis that proteins which are incapable of editing are less toxic to the organism overexpressing them. However, the difference observed here could have been due to several different reasons. The same amount of DNA was used to transform wild type and mutant constructs however there may have been a difference in transformation efficiency. Variation may have been introduced during induction or growth due to starting culture differences. Therefore this observation was noted and not pursued as purified protein was obtained in sufficient amounts for biochemical analysis.

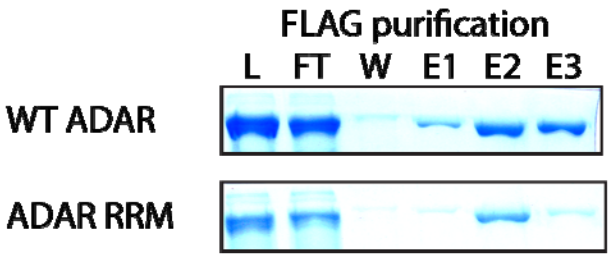


Figure AI.1: Purification of WT ADAR and ADAR RRM from *Pichia pastoris*

Pichia pastoris were transformed with constructs containing WT ADAR or ADAR RRM, expression was induced with methanol and purification was performed as described in Ring *et al*, (2004). Aliquots were taken at each stage of the purification for analysis by coomassie-stained gel, the FLAG purification steps are shown. An equal volume of 10ul was loaded for each fraction of WT and RRM mutant protein preparations. Where L is the load fraction, FT is the flow through, W is the wash, and E1-E3 are elution fractions. More protein is present in the WT ADAR sample at all stages of purification.

However, the same mutations (KKxxK->EAxxA) were subcloned under a metallothionein promoter for use in *Drosophila* S2 cell culture. $1x10^6$ S2 cells were plated and transfected with 2ug DNA encoding either WT or RRM mutant ADAR. Following induction for 24 hours with 0.6mM copper sulphate cells were harvested. Protein content was determined by Bradford assay and 60ug total protein was loaded onto a PAGE gel. Immunoblot analysis with an anti-FLAG antibody revealed that there was a higher amount of WT ADAR protein than ADAR RRM protein (Figure AI.2). Both WT and RRM-mutant ADAR migrate at approximately 95kDa. A non-specific protein of approximately 60kDa is detected with the anti-FLAG antibody which indicates that equal amounts of protein were loaded in all lanes. Both WT ADAR and ADAR RRM were only present after induction with copper sulphate. Immunocytochemistry revealed that both wild type ADAR and ADAR RRM had nuclear localisation (data not shown).

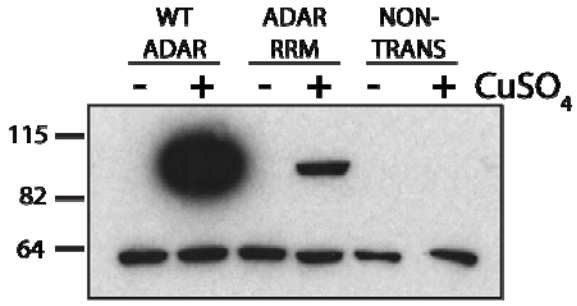


Figure AI.2: *Drosophila* S2 cell expression of WT ADAR and ADAR RRM
$1x10^6$ *Drosophila* S2 cells were transfected with cellfectin reagent (Invitrogen) and 2ug plasmid DNA encoding either WT ADAR or ADAR RRM, or non-transfected as a control. Protein expression was induced in half the samples by addition of 0.6mM copper sulphate (indicated by +/-). Extracts were quatified using a Bradford assay and 60ug protein was loaded per lane. Immunoblot analysis was performed using an anti-FLAG antibody. Considerably higher amounts of WT ADAR protein are produced than ADAR RRM, although quantification is not possible from this immunoblot. A non-specific band detected by the anti-FLAG antibody indicates loading was equal for lanes containing ADAR proteins.

Individually the observations shown here are not remarkable as they could be explained by variation in experimental procedure. However, when analysed together with the observation that ADAR EA protein is consistently expressed to a higher levels than WT ADAR they indicate that the ADAR RRM protein may not be as stable as wild type protein. This is further supported by the observation that an ADAR2-RRM protein expressed in HEK293T cells is also found at a lower level than other ADAR2 mutants when cells are transfected and treated in an identical manner (Roberto Marcucci, unpublished observation). This indicates that this phenomenon is independent of the expression system used as it has been observed in three independent systems; yeast, *Drosophila* S2 cells, and HEK293T cells. It also indicates it is independent of the species of origin the ADAR used as the same observations are made with *Drosophila* ADAR and human ADAR2. Further, it is unlikely to be a problem with the process of site-directed mutagenesis as different primers were required for *Drosophila* ADAR and human ADAR2, and constructs were sequenced in their entirety prior to use (Jim Brindle). However investigation of WT ADAR and ADAR RRM by circular dichroism is required to determine whether the mutant protein has altered stability.

These mutants were generated and used in this thesis due to criticism of previous ADAR RNA-binding mutants generated in the lab (Gallo et al., 2003) on the basis that they disrupted the protein structure (Valente and Nishikura, 2007). The observations presented here indicate that this RNA-binding mutation (KKxxK->EAxxA) creates a protein that is either unstable or unfavourable to the organism expressing it. However, in the initial paper describing these mutants size exclusion chromatography was performed on a gel filtration column which indicates stable dimers were formed. Therefore it cannot be excluded that these mutations generate a gain of function which is undesirable to the host organism. *In vitro* analysis indicates that RNA-binding is impaired in this mutant (Figure 4.9), and no editing of a known substrate was observed (Figure 4.13).

The implications of this observation are that the flies generated to analyse the human ADAR2-RRM mutant may not be expressing the protein to a high level. This is

surprising as the neurodegeneration observed in these flies is not as advanced as that seen in age-matched *Adar*^{5G1} flies, although it is clearly present. However, due to the low expression of ADAR protein in flies it would be very difficult to measure it.

Chapter 9: Appendix II

This Appendix details the gene lists obtained from microarray analysis described in Chapters 4 & 5.

Details of primers used are listed in Table AII.4

Gene	Accession	w^{1118} mean	$Adar^{5G1}$ mean	Fold change	$Adar^{5G1}$ mean	Adar mean	Fold change	$Adar^{5G1}$ mean	Adar EA mean	Fold change
X67681 /transposon	Transposon.8	123.7	879.1	7.11	879.1	49.12	-17.9	879.1	105.76	-8.31
schnurri	Transposon.24	1784.93	2659.53	1.49	2659.53	1463.64	-1.82	2659.53	1310.97	-2.03
AY180916 /transposon	Transposon.38	2418.3	4453.24	1.84	4453.24	3106.89	-1.43	4453.24	3023.28	-1.47
GH06606 /chr2R:-9953073,9956759	GH06606	299.28	455.34	1.52	455.34	293.08	-1.55	455.34	293.56	-1.55
CG3513	CG3513-RA	219.02	496.43	2.27	496.43	282.67	-1.76	496.43	224.79	-2.21
TpnC25D	CG6514-RA	4185.97	6279.91	1.5	6279.91	3305.83	-1.9	6279.91	4569.09	-1.37
CG5958	CG5958-RA	1537.35	4743.55	3.09	4743.55	3235.73	-1.47	4743.55	3590.9	-1.32
Cyp4d21	CG6730-RA	3943.92	7076.92	1.79	7076.92	4267.98	-1.66	7076.92	4194.09	-1.69
CG31904 /Adult cuticle protein 1	CG7216-RA	227.07	624.99	2.75	624.99	211.02	-2.96	624.99	241.47	-2.59
CG7203	CG7203-RA	2618.27	5337.51	2.04	5337.51	1749.09	-3.05	5337.51	1528.75	-3.49
CG7214	CG7214-RA	57.17	243.99	4.27	243.99	58.21	-4.19	243.99	53.55	-4.56
CG15267	CG15267-RA	294.34	411.39	1.4	411.39	260.25	-1.58	411.39	293.9	-1.4
CG31780 /CG18477	CG31780-RB	97	218.08	2.25	218.08	92.33	-2.36	218.08	59.12	-3.69
CG9317	CG9317-RB	1027.07	1420.22	1.38	1420.22	755.11	-1.88	1420.22	729.72	-1.95
CG14401	CG14401-RA	552.29	824.99	1.49	824.99	576.95	-1.43	824.99	562.93	-1.47
CG6675	CG6675-RA	301.41	479.39	1.59	479.39	260.61	-1.84	479.39	205.47	-2.33
HDC00331 /chr2L:-2185138,2185389	HDC00331	2537.42	3779.65	1.49	3779.65	1973.28	-1.92	3779.65	2086.61	-1.81
Immune induced molecule 1	CG18108-RA	6011.64	11153.7	1.86	11153.7	6862.32	-1.63	11153.7	5818.25	-1.92
Glutathione S transferase E1	CG5164-RA	2723.92	4083.45	1.5	4083.45	2650.2	-1.54	4083.45	2493.17	-1.64
Immune induced molecule 23	CG15066-RA	1643.33	7130.36	4.34	7130.36	2764.76	-2.58	7130.36	2559.28	-2.79
CG30091	CG30091-RA	86.97	441.73	5.08	441.73	80.25	-5.5	441.73	22.62	-19.53
CG30050	CG30050-RA	242.45	510.29	2.1	510.29	264.86	-1.93	510.29	344.82	-1.48
Troponin C at 47D	CG9073-RA	68.84	406.43	5.9	406.43	154.11	-2.64	406.43	102.59	-3.96
FMRFamide-related	CG2346-RA	415.48	618.87	1.49	618.87	356.68	-1.74	618.87	450.09	-1.37

Tetraspanin 42Ei	CG12843-RA	977.4	1473.75	1.51	1473.75	1087.45	-1.36	1473.75	1084.06	-1.36
CG15666	CG15666-RA	243.76	381.05	1.56	381.05	209.98	-1.81	381.05	193.24	-1.97
maleless	CG11680-RB	437.13	693.05	1.59	693.05	352.48	-1.97	693.05	337.21	-2.06
Cytochrome P450-6a2	CG9438-RA	4149.61	5772.75	1.39	5772.75	3925.73	-1.47	5772.75	3636.01	-1.59
CG12858	CG12858-RA	1274.05	1938.64	1.52	1938.64	939.28	-2.06	1938.64	1125.05	-1.72
CG2065	CG2065-RA	311.16	983.51	3.16	983.51	341.47	-2.88	983.51	253.51	-3.88
SCAP	CG33131-RA	656.26	1268.66	1.93	1268.66	848.83	-1.49	1268.66	932.27	-1.36
Sarcoplasmic calcium-binding protein 1	CG15848-RA	1530.1	3495.9	2.28	3495.9	913.64	-3.83	3495.9	1032.18	-3.39
CG8789	CG8789-RA	586.42	885.36	1.51	885.36	606.76	-1.46	885.36	659.72	-1.34
CG6812	CG6812-RA	1198.8	1616	1.35	1616	1204.46	-1.34	1616	1194.14	-1.35
Neuropeptide-like precursor 3 /Nplp3	CG13061-RA	10303.13	15591.2	1.51	15591.2	9575.77	-1.63	15591.2	7086.57	-2.2
CG17029	CG17029-RA	757.56	1257.32	1.66	1257.32	855.8	-1.47	1257.32	750.55	-1.68
CG17364	CG17364-RB	29.55	170.19	5.76	170.19	23.93	-7.11	170.19	23.9	-7.12
CG6673	CG6673-RA	568.34	895.53	1.58	895.53	429.37	-2.09	895.53	350.3	-2.56
CG8012	CG8012-RA	923.83	2132.21	2.31	2132.21	1070.13	-1.99	2132.21	970.18	-2.2
Lcp65Ag2	CG10534-RA	212.04	347.48	1.64	347.48	182.03	-1.91	347.48	153.91	-2.26
methuselah-like 8	CG32475-RA	43.75	175.49	4.01	175.49	44.24	-3.97	175.49	52.65	-3.33
methuselah-like 8	CT37020	24.6	226.33	9.2	226.33	27.16	-8.33	226.33	42.65	-5.31
CG10999	CG10999-RA	423.93	604.22	1.43	604.22	456.55	-1.32	604.22	426.75	-1.42
CG1090	CG1090-RA	306.27	538.51	1.76	538.51	395.01	-1.36	538.51	402.93	-1.34
Pheromone-binding protein-related protein 4	CG1176-RA	1741.72	2443.03	1.4	2443.03	1647.08	-1.48	2443.03	1700.25	-1.44
CG12256	CG12256-RA	133.65	261.92	1.96	261.92	154.3	-1.7	261.92	108.94	-2.4
CG14298 /GA12887	CG14298-RA	456.49	645.31	1.41	645.31	408.69	-1.58	645.31	351.21	-1.84
CG5835 /CG11779	CG5835-RA	869.69	1268.03	1.46	1268.03	824.83	-1.54	1268.03	615.56	-2.06
CG13833	CG13833-RA	3765.67	6617.96	1.76	6617.96	4733.5	-1.4	6617.96	3727.88	-1.78

CG13822	CG13822-RA	104.09	781.7	7.51	781.7	419.48	-1.86	781.7	495.48	-1.58
CG33970	CG14559-RA	407.7	553.82	1.36	553.82	393.51	-1.41	553.82	344.83	-1.61
CG33970	CG31085-RB	3155.79	4198.03	1.33	4198.03	3082.43	-1.36	4198.03	2816.36	-1.49
CG12054	CG12054-RA	489.14	853.65	1.75	853.65	532.52	-1.6	853.65	608.46	-1.4
Metallothionein D	CG33192-RA	128.51	564.56	4.39	564.56	91.44	-6.17	564.56	155.94	-3.62
lethal (1) G0237	CG1558-RA	502.68	755.5	1.5	755.5	386.29	-1.96	755.5	516.71	-1.46
furrowed	CG1500-RA	287.11	491.6	1.71	491.6	310.51	-1.58	491.6	288.86	-1.7
found in neurons	CG4396-RA	681.95	966.96	1.42	966.96	598.4	-1.62	966.96	710.19	-1.36
Cyp6t1	CG1644-RA	704.68	1110.21	1.58	1110.21	833.96	-1.33	1110.21	520.18	-2.13
CG40486	CG40486-RB	3040.75	4591.72	1.51	4591.72	2812.01	-1.63	4591.72	3021.03	-1.52

Table AII.1: Genes which are upregulated in *Adar*^{5G1} flies when compared to *w*^{1118} and restored by expression of an active *Adar* or catalytically inactive *Adar EA* transgene in the cholinergic neurons.

The data is shown as a heat map in Figure 4.2. The mean expression value is given for each gene and the fold change when comparing samples in a pairwise fashion.

Gene	Accession	w^{1118} mean	$Adar^{5G1}$ mean	Fold change	$Adar^{5G1}$ mean	Adar mean	Fold change	$Adar^{5G1}$ mean	Adar EA mean	Fold change
X01472 /transposon	Transposon.2	3553.61	1577.1	-2.25	1577.1	5854.68	3.71	1577.1	5637.09	3.57
Z27119 /transposon	Transposon.35	1623.81	330.91	-4.91	330.91	546.78	1.65	330.91	656.16	1.98
AF365402 /transposon	Transposon.47	600.85	260.69	-2.3	260.69	468.93	1.8	260.69	509.95	1.96
HDC11981 /chrXh:-143303,143549	HDC11981	2641.87	308.74	-8.56	308.74	856.85	2.78	308.74	1047.72	3.39
LP09838 /chr2h:-742011,743734	LP09838	549.89	174.37	-3.15	174.37	277.36	1.59	174.37	325.95	1.87
hoepel2	CG15624-RA	1032.33	457.55	-2.26	457.55	759.18	1.66	457.55	676.49	1.48
CG31955	CG31955-RA	552.78	329.67	-1.68	329.67	597.22	1.81	329.67	564.03	1.71
CG17646	CG17646-RB	3065.04	2144.51	-1.43	2144.51	3252.47	1.52	2144.51	3312.56	1.54
antennal protein 5	CG5430-RA	1991.57	1104.75	-1.8	1104.75	1542.15	1.4	1104.75	1865.47	1.69
CG5156	CG5156-RA	361.58	148.21	-2.44	148.21	554.08	3.74	148.21	360.22	2.43
Myo28B1	CG6976-RC	294.39	190.46	-1.55	190.46	415.26	2.18	190.46	390.4	2.05
lectin-28C	CG7106-RA	2909.72	1646.08	-1.77	1646.08	4504.01	2.74	1646.08	3098.63	1.88
CG4364	CG4364-RA	2852.78	1922.9	-1.48	1922.9	2851.63	1.48	1922.9	3007.21	1.56
CG6583	CG6583-RA	1574.02	911.18	-1.73	911.18	1304.39	1.43	911.18	1254.37	1.38
vasa intronic gene	CG4170-RD	2333.61	1556.1	-1.5	1556.1	2139.05	1.37	1556.1	2218.86	1.43
dumpy	CG33196-RB	510.33	304.62	-1.68	304.62	508.63	1.67	304.62	518.03	1.7
G protein 30A	CT12391	917.3	469.09	-1.96	469.09	737.05	1.57	469.09	681.29	1.45
CG33679	Dm.2L.3922.0	177.62	56.69	-3.13	56.69	234.7	4.14	56.69	167.77	2.96
CG11474	CG11474-RB	778.45	435.53	-1.79	435.53	742.99	1.71	435.53	675.64	1.55
Cyp6d2	CG4373-RA	432.59	255.84	-1.69	255.84	426.26	1.67	255.84	386.13	1.51
CG3105	CG3105-RA	600.05	444.8	-1.35	444.8	595.23	1.34	444.8	633.31	1.42
RpL12 /Ribosomal protein L12	CG3195-RC	604.64	416.38	-1.45	416.38	675.88	1.62	416.38	658.83	1.58
Cytochrome P450-9c1	CG3616-RA	516.75	269.2	-1.92	269.2	668.06	2.48	269.2	447.92	1.66
Glutathione S transferase E5	CG17527-RA	809.01	295.88	-2.73	295.88	748.93	2.53	295.88	602.6	2.04
Glutathione S transferase E7	CG17531-RA	689.46	430.53	-1.6	430.53	1151.74	2.68	430.53	939.75	2.18
CG30118	CG30118-RA	1897.82	1328.48	-1.43	1328.48	1964.43	1.48	1328.48	2163.72	1.63

CG30083	CG30083-RA	514.98	304.86	-1.69	304.86	1382.64	4.54	304.86	1189.67	3.9
Pyruvate dehydrogenase kinase	CG8808-RA	3791.99	2426.02	-1.56	2426.02	3650.23	1.5	2426.02	4699.27	1.94
Cyp4p2	CG1944-RA	505.62	249.79	-2.02	249.79	643.68	2.58	249.79	674.02	2.7
Cyp4p3	CG10843-RA	296.42	179.86	-1.65	179.86	304.27	1.69	179.86	343.23	1.91
photorepair	CG11205-RA	710.6	48.39	-14.68	48.39	437.31	9.04	48.39	458.64	9.48
CG18853 / photorepair	CG11205-RA	1342.92	86.2	-15.58	86.2	669.44	7.77	86.2	669.31	7.77
CG30382 /Proteasome 6 subunit	CG30382-RA	3892.59	2193.29	-1.77	2193.29	3053.41	1.39	2193.29	2876	1.31
Tetraspanin 42Eg	CG12142-RA	507.92	330.95	-1.53	330.95	575.51	1.74	330.95	573.73	1.73
Spn43Ad	CG1859-RA	322.67	208.19	-1.55	208.19	517.86	2.49	208.19	423.47	2.03
CG30404	CG30404-RA	1693.1	972.71	-1.74	972.71	1373.66	1.41	972.71	1412.15	1.45
CG3270	CG3270-RA	1067.4	310.25	-3.44	310.25	882.66	2.84	310.25	646.67	2.08
CG11400	CG11400-RA	2324.44	1645.46	-1.41	1645.46	3264.25	1.98	1645.46	2963.92	1.8
Cyp6a20	CG10245-RA	1832.47	610.98	-3	610.98	1932.1	3.16	610.98	2424.86	3.97
Argonaute 1	CG6671-RC	1031.55	653.15	-1.58	653.15	862.25	1.32	653.15	889.68	1.36
CG2064	CG2064-RA	904.4	384.63	-2.35	384.63	874.43	2.27	384.63	828.11	2.15
CG30492	CG30492-RE	1563.93	839.8	-1.86	839.8	1678	2	839.8	1362.54	1.62
CG8801	CG8801-RA	4039.11	2945.09	-1.37	2945.09	4959.19	1.68	2945.09	5454.7	1.85
RH09485 /chr2R:+13775110,13775672	RH09485	2412.16	1376.53	-1.75	1376.53	2812.35	2.04	1376.53	2120.22	1.54
CG13877	CG13877-RA	1700.88	642.6	-2.65	642.6	970.7	1.51	642.6	1188.69	1.85
CG13704	CG13704-RA	596.75	269.94	-2.21	269.94	1959.14	7.26	269.94	1535.74	5.69
CG18135	CG18135-RB	7334.06	4317.52	-1.7	4317.52	9491.81	2.2	4317.52	10787.1	2.5
lethal (3) neo26 /l(3)neo26	CG6874-RA	347.81	213.73	-1.63	213.73	386.98	1.81	213.73	394.65	1.85
ypsilon schachtel	CG5654-RA	6555	4385.02	-1.49	4385.02	5907.92	1.35	4385.02	6629.55	1.51
insulin-like peptide 3	CG14167-RA	518.61	207.65	-2.5	207.65	494.99	2.38	207.65	370.89	1.79
CG3552	CG3552-RA	384.89	242.28	-1.59	242.28	356.98	1.47	242.28	360.94	1.49
CG5741	CG5741-RA	599.7	246.98	-2.43	246.98	458.34	1.86	246.98	397.62	1.61
liquid facets	CG8532-RB	2648.37	1926.28	-1.37	1926.28	2538.05	1.32	1926.28	2708.1	1.41
CG33045-RA	CG33045-RA	2838.03	1882.46	-1.51	1882.46	3570.25	1.9	1882.46	3888.01	2.07

/Kaz1/synonyms:(CG1220)										
CG1315	CG1315-RA	295.88	128.67	-2.3	128.67	362.23	2.82	128.67	277.5	2.16
CG31349/pyd /synonyms:(CG9763 CG11782 CG9729 CG12409 CG11962)	CG31349-RE	899.67	575.89	-1.56	575.89	951.37	1.65	575.89	926.04	1.61
CG9363	CG9363-RB	1108.13	754.81	-1.47	754.81	1383.64	1.83	754.81	1336.7	1.77
CG8147	CG8147-RA	1216.76	424.78	-2.86	424.78	2084.3	4.91	424.78	3482.18	8.2
fau	CG6544-RA	756.31	546.56	-1.38	546.56	767.64	1.4	546.56	839.35	1.54
nocturnin	CG31299-RB	2002.69	1387.28	-1.44	1387.28	2189.05	1.58	1387.28	2057.22	1.48
CG6764	CG6764-RA	4169.88	3064.69	-1.36	3064.69	5076.14	1.66	3064.69	5705.87	1.86
CG17227	CG17227-RA	532.02	275.84	-1.93	275.84	515.38	1.87	275.84	449.51	1.63
CG31469	CG31469-RA	467.39	293.31	-1.59	293.31	499.24	1.7	293.31	454.51	1.55
CG12333	CG12333-RA	761.14	561.47	-1.36	561.47	780.61	1.39	561.47	776.09	1.38
Cyp12a4	CG6042-RA	2438.43	964.08	-2.53	964.08	2285.64	2.37	964.08	1944.52	2.02
CG6040	CG6040-RA	693.38	500.37	-1.39	500.37	738.71	1.48	500.37	677.39	1.35
alternative testis transcripts ORF B/ORF A	CG4241-RA	668.43	487.28	-1.37	487.28	716.64	1.47	487.28	819.98	1.68
CG6656	CG6656-RA	3092.07	1969.48	-1.57	1969.48	3102.57	1.58	1969.48	3229.56	1.64
CG13641	CG13641-RA	515.48	122.39	-4.21	122.39	266.96	2.18	122.39	229.09	1.87
CG12656	CG12656-RA	2976.86	2115.03	-1.41	2115.03	3058.46	1.45	2115.03	2815.54	1.33
CG3603	CG3603-RA	874.07	372.41	-2.35	372.41	629.61	1.69	372.41	617.15	1.66
CG14419	CG14419-RA	731.05	472.35	-1.55	472.35	870.56	1.84	472.35	689.4	1.46
CG15784	CG15784-RA	1659.61	1002.95	-1.65	1002.95	1409.67	1.41	1002.95	1544.86	1.54

Table AII.1: Genes which are downregulated in *Adar*^{5G1} flies when compared to *w*^{1118} and restored by expression of an active *Adar* or catalytically inactive *Adar EA* transgene in the cholinergic neurons.

The data is shown as a heat map in Figure 4.3. The mean expression value is given for each gene and the fold change when comparing samples in a pairwise fashion.

Table AII.3: 759 genes which were significantly altered according to AS array analysis of w^{1118} and $Adar^{5G1}$ head RNA.

1	dp	51	GluRIIB	101	Tsp42Ea	151	MAPk-Ak2
2	Strn-Mlck	52	CG42513	102	CG5973	152	cnn
3	syd	53	Top1	103	CG5594	153	jim
4	ord	54	CG1244	104	Gprk1	154	Sox21b
5	CG13900	55	gig	105	CG34404	155	CG30296
6	Adh	56	Ank2	106	sqd	156	Fmrf
7	TepIII	57	cora	107	Aldh-III	157	CG8974
8	CG31190	58	Phm	108	Atg18	158	vkg
9	Sap47	59	n-syb	109	CG11147	159	Octbeta3R
10	CG3734	60	CG2258	110	CG11293	160	CG9312
11	Cyp4p1	61	tws	111	Gr64d	161	l(2)37Cc
12	mod(mdg4)	62	CG8777	112	CkIIbeta	162	CG5224
13	GRHR	63	tacc	113	CG9698	163	GluRIIE
14	CG5835	64	RhoGEF2	114	Trp1	164	CG10737
15	Mf	65	CG14969	115	CG42249	165	fau
16	BRWD3	66	Pten	116	ndl	166	su(Hw)
17	CG3999	67	cn	117	CG17549	167	GS
18	CG31660	68	CG6876	118	CG11347	168	CG10184
19	CG10359	69	CG14782	119	Wwox	169	Pmi
20	CLIP-190	70	Pk92B	120	CG10418	170	Ppn
21	CG42541	71	Mhc	121	RhoGAP18B	171	AnnIX
22	CG30463	72	Hsf	122	mask	172	CG12090
23	wupA	73	Ca-P60A	123	CG3566	173	CG5261
24	CG31352	74	CG10137	124	CG3160	174	14-3-3zeta
25	heph	75	shep	125	Gycalpha99B	175	unc-104
26	B52	76	CG8945	126	Dp	176	eIF2B-alpha
27	Dhc36C	77	Gr64e	127	CG33521	177	TepII
28	Fmo-2	78	CG31033	128	CG10186	178	CG30424
29	CG14480	79	slo	129	CG1998	179	SK
30	cbs	80	l(1)G0255	130	CG10249	180	CG12370
31	unc-115	81	CG15553	131	zormin	181	CG42314
32	syt	82	kay	132	CG14864	182	smp-30
33	lola	83	CG8422	133	Nipped-B	183	trpl
34	cher	84	Cyp4ac3	134	MESK2	184	hang
35	grh	85	CkIIalpha-i3	135	Dys	185	UBL3
36	CG9817	86	CG13168	136	Bsg	186	SP2637
37	CG4662	87	CG3437	137	stmA	187	Sxl
38	dpr10	88	CG7741	138	CG7059	188	CtBP
39	Unc-89	89	CG8399	139	Pax	189	CG31826
40	Cpr47Ef	90	CG9812	140	tomosyn	190	X11Lbeta
41	Chro	91	CG31469	141	Fas2	191	CG17928
42	GluRIIC	92	scaf6	142	Cam	192	l(2)gl
43	CG5335	93	olf186-F	143	Rbf	193	CG8336
44	Tsp29Fb	94	CG12384	144	CG1826	194	Adhr
45	para	95	CG9855	145	CG17471	195	CG11951
46	beat-VII	96	CG10631	146	Galpha49B	196	l(1)G0193
47	Nmdar1	97	CG5621	147	opa1-like	197	fbp
48	CG6043	98	unc-13	148	Lip1	198	CG11236
49	CG5895	99	Eph	149	CG7945	199	CG3153
50	bw	100	CG6206	150	CG3792	200	CG31158

201	CG33087	251	CG32169	301	Pkc53E	351	CG11961
202	CG33722	252	Lmpt	302	HDAC4	352	poe
203	CG32391	253	CG7145	303	CG32687	353	CG5815
204	Argk	254	Ranbp16	304	CG1275	354	Nup44A
205	CG34114	255	brp	305	CycA	355	CG3711
206	CG5080	256	eIF3-S8	306	CG2100	356	Rala
207	l(1)G0232	257	btv	307	Axs	357	dpr8
208	Gycbeta100B	258	CG7597	308	CG6830	358	Moca-cyp
209	CG9649	259	CG5009	309	l(2)37Ce	359	Mcm7
210	Lsd-1	260	pUf68	310	psh	360	trx
211	Pi3K68D	261	CG4069	311	stai	361	CG8108
212	nAcRbeta-64B	262	Rap2l	312	spir	362	CG32082
213	Graf	263	CG1969	313	Rgk2	363	CG8086
214	CG7695	264	Rab27	314	tra2	364	exo84
215	Cyp313a1	265	CG12531	315	CG16838	365	CG14476
216	Prm	266	CG31012	316	ced-6	366	CG1792
217	CG3321	267	tlk	317	Cbp53E	367	CG6123
218	CG11474	268	Sodh-2	318	CG15111	368	CG1041
219	fwd	269	CG11306	319	CG4301	369	inaD
220	CG17150	270	bru-2	320	CG8963	370	mRpS9
221	Stim	271	trol	321	CG18304	371	scrib
222	sws	272	Tango7	322	CG9510	372	CG17494
223	Trl	273	Moe	323	PNUTS	373	CG11843
224	Fbxl4	274	Mlc1	324	CG5445	374	Dhc93AB
225	CG6928	275	CG31716	325	CG32808	375	CG5946
226	Ets65A	276	CG9634	326	cact	376	Mical
227	CG4829	277	CG3703	327	CG8709	377	CG3831
228	CG30197	278	lqf	328	mRNA-capping-enzyme	378	Or45b
229	CG8216	279	RpS9	329	CG8475	379	CG2478
230	Atg6	280	CG12184	330	SP2353	380	CG12766
231	yellow-e	281	Ptp4E	331	CG5060	381	sw
232	ia2	282	Oscillin	332	CG9628	382	sif
233	Myo28B1	283	CG8083	333	Asph	383	alphaTub84B
234	glo	284	CG30492	334	gfA	384	AP-1gamma
235	Caps	285	Msr-110	335	CG18749	385	CPTI
236	Gel	286	sec71	336	PGRP-LD	386	CG4406
237	jar	287	CG11875	337	CG10254	387	CG7971
238	CG33107	288	CG3493	338	mnd	388	CG17333
239	CG34400	289	Pbprp3	339	alien	389	CG5392
240	NijA	290	Zasp	340	CG2233	390	CG1208
241	Top3beta	291	Pbprp1	341	CG17665	391	CG4587
242	Rim	292	eIF-2beta	342	Cyp28a5	392	CG3609
243	nocturnin	293	jbug	343	CG14204	393	AP-1sigma
244	tna	294	CG15373	344	CG6842	394	Dgkepsilon
245	CG9196	295	CG6114	345	CG10621	395	Treh
246	e(r)	296	CG14935	346	gammaSnap	396	mus210
247	CG6084	297	CG12340	347	dpr15	397	CG1814
248	CG6891	298	CG3907	348	Baldspot	398	CG11155
249	Aats-asp	299	LanB1	349	Spn4	399	CG13551
250	CG17082	300	GABA-B-R1	350	Pde6	400	CG33097

401	Spt-I	451	Tsp42Eq	501	inx3	551	CG7698
402	Tm2	452	stops	502	Spn1	552	Crk
403	CG4019	453	alpha-Spec	503	mbl	553	CG8632
404	bgm	454	CG7927	504	CG3654	554	Nap1
405	pho	455	Rgk1	505	PGRP-LC	555	CG11210
406	nrm	456	Hml	506	alpha-Man-I	556	brat
407	CG10238	457	CG32016	507	CG14215	557	ImpL2
408	MED15	458	CG11486	508	pck	558	lok
409	RpS15	459	CG9027	509	lig	559	CG13913
410	mab-21	460	ERR	510	qkr58E-1	560	ImpL3
411	CG15626	461	XRCC1	511	RhoGAP19D	561	CG33056
412	Tre	462	CG14207	512	Tcp-1eta	562	CG10664
413	Pkcdelta	463	Fur2	513	CG12206	563	Pi4KIIalpha
414	cv	464	boca	514	Liprin-alpha	564	CG15434
415	RpL13	465	CG8798	515	jigr1	565	Pros29
416	CG6287	466	CG10019	516	CG4660	566	CG5708
417	CG30427	467	CG17295	517	Dscam	567	PP2A-B'
418	tsl	468	CG1942	518	CG5955	568	SmG
419	Jhl-26	469	ena	519	miple2	569	RpL7A
420	rab3-GEF	470	CG14691	520	nAcRalpha-96Aa	570	CG5853
421	CG4045	471	CG17698	521	rdgA	571	CG11876
422	Dyrk3	472	RpLP0	522	CG18616	572	MRP
423	tho2	473	hts	523	glob1	573	CG41476
424	CG6199	474	CG7920	524	CG7470	574	CG17751
425	CG3045	475	CG2604	525	CG11069	575	CG1909
426	NtR	476	CG11796	526	toy	576	Vmat
427	ACXC	477	CG5524	527	CG30116	577	CG7255
428	RluA-1	478	Tpi	528	CG4678	578	Srp14
429	eIF-5A	479	Syt7	529	CG32103	579	CG31678
430	CG7382	480	lh	530	ppk13	580	fry
431	CG5549	481	CG3074	531	CaMKII	581	CG31974
432	CG3585	482	CG15390	532	CG8788	582	gish
433	sev	483	CG15478	533	CG1688	583	CG4673
434	CG7272	484	Vha68-1	534	Cyp9f2	584	key
435	CG2750	485	CG4872	535	na	585	CG8234
436	CG12006	486	CG32495	536	CYLD	586	Rbp1
437	CG5793	487	CG4000	537	RpL23	587	ana1
438	zip	488	CG9919	538	RpS7	588	pip
439	CG8121	489	Mkp	539	CG7888	589	Ance-4
440	CG14585	490	CG42271	540	CG9588	590	CG4221
441	CG31823	491	CG5945	541	alpha-Cat	591	CG6643
442	CG4658	492	Smr	542	CG15449	592	CG5541
443	coilin	493	E2f2	543	CG10592	593	cpx
444	form3	494	CG11163	544	CG6912	594	CG17816
445	CG40470	495	Cchl	545	CG42342	595	CG7834
446	CG31226	496	cals	546	Jupiter	596	CG8199
447	SmB	497	CG30359	547	GlcAT-P	597	Nipsnap
448	gry	498	pan	548	CG32210	598	Msp-300
449	CG32626	499	Nedd4	549	Syn	599	smi35A
450	CG11779	500	CG5527	550	Caki	600	CalpA

601	CG5770	651	CG4820	701	CG13698	751	sesB
602	CG13868	652	CG4757	702	CG8129	752	grn
603	CG31619	653	CG33275	703	CG17233	753	CG32176
604	CG9279	654	ldgf4	704	CG8335	754	pug
605	fbl	655	CG10237	705	Pkn	755	atl
606	tow	656	CG30096	706	Pbprp2	756	CG5482
607	chic	657	CG32240	707	ND75	757	CG11859
608	CG17352	658	eIF4G	708	RpL13A	758	mtacp1
609	CG31183	659	Ance	709	Tm1	759	CG8334
610	CG15117	660	vib	710	CG6749		
611	CG7828	661	CG14740	711	grass		
612	CG18003	662	gcl	712	lap2		
613	Fer1HCH	663	pum	713	CG1518		
614	CG8043	664	Fancd2	714	CG15539		
615	Coq2	665	Cyp4c3	715	Dredd		
616	CG13384	666	CG16718	716	CG18789		
617	CG13409	667	TFAM	717	l(2)05510		
618	Mnt	668	CG17378	718	Pvr		
619	CG6357	669	G-salpha60A	719	pr-set7		
620	Menl-2	670	Tl	720	CG30345		
621	StlP	671	CG8646	721	Nc73EF		
622	CG7646	672	CG8485	722	bl		
623	CG31755	673	CG1091	723	CG4849		
624	Mhcl	674	amon	724	Pkg21D		
625	armi	675	alt	725	pHCl		
626	CG9455	676	Rya-r44F	726	CG6129		
627	CG11652	677	T48	727	CG34363		
628	CG9222	678	phtf	728	PGRP-SB1		
629	snama	679	CAP	729	CG13397		
630	CG15312	680	MTF-1	730	CG15096		
631	mRpL36	681	Nmdar2	731	dlg1		
632	slgA	682	CG32364	732	CG7447		
633	Cp1	683	RpL11	733	Mapmodulin		
634	CG32432	684	Rab14	734	CG15093		
635	Lgr3	685	CG13876	735	Sb		
636	Mkp3	686	CG3209	736	CG18788		
637	MTA1-like	687	Atg7	737	bi		
638	CG6424	688	TER94	738	CG3815		
639	CG15564	689	Adar	739	pAbp		
640	bip2	690	CG9896	740	CG6465		
641	CG4004	691	Tip60	741	CG3884		
642	CG8176	692	CG1139	742	Atg5		
643	Trip1	693	fs(2)ltoPP43	743	RpL35A		
644	eIF-4a	694	Not1	744	CG9448		
645	CG11148	695	rg	745	Dek		
646	CG42390	696	CG7017	746	CG15347		
647	AP-2sigma	697	PH4alphaNE3	747	CG7601		
648	Klp31E	698	CG11550	748	SamDC		
649	CG31116	699	mspo	749	CG8320		
650	ninaG	700	CG1824	750	Mbs		

Table AII.4: Primer sequences.

Name	Sequence 5'-3'
Mouse expression primers	
Apo4_5'UTR_For	GGAGACAAAGATGATTTGCAGC
Apo4_3'UTR_Rev	CTAGAGTCTGTTTCTTCCATTGGC
GAPDH_For	ACCACAGTCCATGCCATCAC
GAPDH_Rev	TCCACCACCCTGTTGCTGT
Primers for cloning A4	
SpeI_Apo4_46aaF	GACAAACTAGTAACAGAATTTCATCAGACT
NdeI_Apo4_F	TGCCATATGGAGCCCCTGTATGAGGAGATT
Apo4_EcoRI_R	GCAGAATTCCCTCTCCAGTAAACGGCAAATTC
***Drosophila* primers**	
Primers for real time PCR	
Atx2_F	AGAACAAGGGCGGCTACC
Atx2_R	CGTTCGAATCGCCATTTAG
Adar_For	CCTAAATGATTCCCATGCTGA
Adar_Rev	ACGAAAATCGACTGATATGCTG
MetallothioneinD_F	GCAAGGCTTGTGGAACAAA
MetallothioneinD_R	TCCGTTCTAGCAGGAGCACT
DILP3_3'_For2	TCCCTGCTGGAAAGACTGTT
DILP3_3'_Rev2	GGCTTGGCAGCACAATATCT
CG30091_For2	GGAGCAGCAGAAGTTCCAAA
CG30091_Rev2	CCTCCCGGTAGCACTTACTG
Phr5'region_For2	GGCGCGTACAGGACAACT
Phr5'region_Rev2	ATTCCAGCTTGAGAGCCAGA
Phr3'UTR_For2	GCTGCATGTGGTCCATTG
Phr3'UTR_Rev2	ACCTAACCTTACCGAAGATAGCC
Rab3-GEF_For	CAGAAGGGTGGCATTTTTGT
Rab3-GEF_Rev	ACCGAGTAGCAAATCTGATCG
Rab3-GEF_3'UTR_F2	TGTTAGCTGGCTCTGGATGTT
Rab3-GEF_3'UTR_R2	TCTCGCTTTGAACTTTGAAAAA
Rdl_3'UTR_F	AAACGAAGCGGAAACAAAAC
Rdl_3'UTR_R	TGTCCTTCCCGCAATTGTAT
Nic34E_3'UTR_F	TTGCAGCTATGGTCGTTGAC
Nic34E_3'UTR_R	TGGCTAATATTGTGAACATTGTGAA
Nic30D_3'UTR_F	CGATTATTGCAACGGTTACG
Nic30D_3'UTR_R	AGCGTCCTTATTGCACGATT
Nic64B_A_3'UTR_F	GCAAACAAACGTTTCCCAAA
Nic64B_A_3'UTR_R	GCGTAAAACTGCGATTACTTATGA
Nic64B_B_3'UTR_F	AACGTTTCGTGTATATAGCTTAGCC
Nic64B_B_3'UTR_R	TGTTGTTGCTGGTCTATTTGCT
Gapdh_For	ACGAGAGTAAAAGTGAAAAGACAGC
Gapdh_Rev	TCCGTTAATTCCGATCTTCG
actin_For	AGGATCGGGATGGTCTTGAT
actin_Rev	CTGCTGCTTCCTCGACTTCT
shaker_For	AACACTAGCGACTGTCGTTGC
shaker_Rev	ACTGGCGCTTTTGGAAGAT
MtnD_3'UTR_F	GCCAGTGCTCCTGCTAGAAC
MtnD_3'UTR_R	GTTGTCATGCCAAAGGGTCT

Name	Sequence 5'-3'
Primers for *in vitro* editing assay	
RP49_5UTR_F	ATTTTGGGCCCACGTGTATT
RP49_3'UTR_R	CATGTTATCAATGGTGCTGCT
RP49_intronF	TGCATTAGTGGGACACCTTG
RP49_intronR	AATTTGGCACAATCCTCGTT
dAdar160_F2	TCGTGAGGAATACTGATGG
dAdar160_R2	TTTGGATGTTTATCAACACCA
Phr_DNA_5'For	GTGTATTTGGAGTGCGGCTA
Phr_DNA_3'Rev	CCCACTTTCCGTATTTTGGA
Phr5'UTR_For	GGCGCGAATCATTCAAGA
Phr3'UTR_Rev	TCTTCTTGTGCACCTTACCG
Phr_582R	CATCAGCAGGTGAAACGGTA
Phr_465F	TTGCCTAGTTCCCAAATTCC
Phr_1209R	GCAAAAGTTATCCGCCAGTT
Phr_1074F	GTGGCTGCACTTTGGTCATA
dILP3_DNA_5'For	GCATCCATACTTAAACACCACTTC
dILP3_DNA_3'UTR	TTGTCGTCATTGGGTTAATAGG
dILP3_5'UTR_For	CATCGAGATGAGGTGTCAGG
dILP3_3'UTR_Rev	GGCAGCACAATATCTCAGCA
Gapdh2_DNA_5'For	GGTAGCATTGAGCTTTCACG
Gapdh2_DNA_3'Rev	GGTATTTCAACGGTTTATACCAA
Alternative splicing analysis primers	
Dscam_jxn14-15_F	GATTCAATAACATTGGAGCT
Dscam_15R	GTTGGACGACACTTCAATG
Dscam_14F	TTCGGAGAATGGGAAACATC
Dscam_14R	TTGAATCCTGTGGCATAGACC
Strn-Mlck_jxn24-25_F	TGATGGAGTACATCACTGGC
Strn-Mlck_25R	ATCGAGATGCACCACACTCT
Strn-Mlck_24F	ACACCGTCAAAAACCTCCAG
Strn-Mlck_24R	CTCCTCTGTGGCGTATTGGT
Strn-Mlck_jxn7-8_F	CCCCTCCGAGATGTGTATCA
Strn-Mlck_8R	ACCACGGGTATCTGGTCCTT
Strn-Mlck_7F	CCCGTAACCTGCACTACCAC
Strn-Mlck_7R	GTCGGCCTTCGACTTAAGGG
Dumpy_32F	ACAGAGTGTCCTGCCAACCT
Dumpy _32Fa	TGTGGCAATGAGGCTATCTG
Strn-Mlck_24Fb	ACCAATACGCCACAGAGGAG
Dscam_14Fb	AGGTCTATGCCACAGGATTCA

Quickchange primers

dsRBD1_EAA_F (with BbvI site)

GGC CAG GGC CGT AGT GAA GCA GTT GCA GCC ATC GAA GCA GCA GCA

dsRBD1_EAA_R

TGC TGC TGC TTC GAT GGC TGC AAC TGC TTC ACT ACG GCC CTG GCC

dsRBD2_EAA_F (with AciI site)

GGA ACA GGT CCT TCC GAA GCG ACG GCC GCA AAT GCG GCA GCT AAG

dsRBD2_EAA_R

CTT AGC TGC CGC ATT TGC GGC CGT CGC TTC GGA AGG ACC TGT TCC